

Daniela Bertolini Depizzol

Uma Abordagem Estatística para o Projeto de Topologias Físicas de Redes Ópticas

Vitória, ES, Brasil

2018

Daniela Bertolini Depizzol

Uma Abordagem Estatística para o Projeto de Topologias Físicas de Redes Ópticas

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Doutor em Engenharia Elétrica.

Programa de Pós-Graduação em Engenharia Elétrica

Centro Tecnológico

Universidade Federal do Espírito Santo

Orientador: Prof. Dr. Marcelo Eduardo Vieira Segatto
Coorientadora: Profa. Dra. Marcia Helena Moreira Paiva

Vitória, ES, Brasil

2018

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

B546a Bertolini Depizzol, Daniela, 1982-
Uma Abordagem Estatística para o Projeto de Topologias Físicas de Redes Ópticas / Daniela Bertolini Depizzol. - 2018. 182 f. : il.

Orientador: Marcelo Eduardo Vieira Segatto.
Coorientadora: Marcia Helena Moreira Paiva.
Tese (Doutorado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Análise de redes (Planejamento). 2. Telecomunicações. 3. Teoria dos grafos. 4. Comunicações ópticas. 5. Interconexão em rede (Telecomunicações). 6. Heurística. I. Vieira Segatto, Marcelo Eduardo. II. Moreira Paiva, Marcia Helena. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

Daniela Bertolini Depizzol

Uma Abordagem Estatística para o Projeto de Topologias Físicas de Redes Ópticas

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Doutor em Engenharia Elétrica.

Aprovada em 26 de novembro de 2018.
Vitória, ES, Brasil.

COMISSÃO EXAMINADORA

Prof. Dr. Marcelo Eduardo Vieira Segatto
Universidade Federal do Espírito Santo - UFES
Orientador

Prof. Dr. Anilton Salles Garcia
Universidade Federal do Espírito Santo - UFES
Examinador Interno

Prof. Dr. Moisés Renato Nunes Ribeiro
Universidade Federal do Espírito Santo - UFES
Examinador Interno

Profa. Dra. Maria Claudia Silva Boeres
Universidade Federal do Espírito Santo - UFES
Examinadora Externa

Prof. Dr. Carmelo José Albanez Bastos Filho
Universidade de Pernambuco - UPE
Examinador Externo

*Dedico este trabalho
ao meu esposo
e ao meu filho.*

Agradecimentos

Agradeço a Deus pela capacidade que nos tem concedido de observar, compreender, interagir e admirar os frutos de Sua criação.

Aos professores orientadores Marcelo Eduardo Vieira Segatto e Marcia Helena Moreira Paiva, pelo apoio e confiança, e pelos valiosos ensinamentos e direcionamentos sem os quais a entrega deste trabalho não seria possível.

Ao meu esposo, Fabio de Oliveira Lima, por além de ser a base do meu suporte familiar (junto com nosso filho Pedrinho), é praticamente um terceiro orientador deste trabalho.

Aos colaboradores: Prof. Dr. Jugurta Montalvão e Dr. Rodrigo Stange Tessinari, que participaram ativamente em momentos importantes no decorrer das pesquisas.

Ao Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) da Universidade Federal do Espírito Santo e seu corpo docente, por todo o esforço em manter um curso de qualidade reconhecida para seus alunos.

A todos os amigos e funcionários do Labtel - Laboratório de Telecomunicações, com os quais o convívio tornava a jornada do doutoramento mais leve.

À Banca Examinadora pela disponibilidade, atenção e colaboração para a conclusão desta tese.

Agradeço finalmente a todos que colaboraram de alguma forma para a realização deste trabalho ou que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigada.

*“Se temer que suspeitem ser sua narrativa
inverídica, lembre-se da probabilidade.”*

John Gay (1685-1732)

Sumário

Sumário	7
Resumo	10
Abstract	11
1 Introdução	12
1.1 Redes ópticas e o número de comprimentos de onda	12
1.2 O projeto de redes ópticas no contexto do presente trabalho	16
1.3 Teoria de grafos para redes ópticas	18
1.4 Invariantes topológicos em redes WDM	21
1.5 Delimitação do tema	23
1.5.1 Objetivo Geral	26
1.5.2 Objetivos Específicos	27
1.6 Contribuições da tese	27
1.7 Organização da tese	29
2 Seleção de atributos para o ranqueamento de invariantes de redes ópticas	30
2.1 As topologias de redes ópticas e os invariantes considerados para a seleção de atributos	32
2.1.1 As redes aleatórias e os invariantes considerados na análise	33
2.1.2 As redes reais analisadas	36
2.2 Seleção de atributos baseada num novo estimador de informação mútua	36
2.3 Resultados gerados a partir do ranqueamento de invariantes	40
2.3.1 Análise de todas as redes de todas as ordens juntas	40
2.3.2 Análise das redes separadas pela ordem n	46
2.3.3 Resposta de λ a alguns invariantes, e de $I(\lambda;k)$ a ordem n da rede	51
2.3.4 Contrastando os resultados com os achados da literatura	53
2.3.5 Observações adicionais em relação ao projeto de topologias de redes ópticas	56
2.4 Conclusão	59

3	Modelagem via funções substitutas para a busca de redes com menores valores de λ	60
3.1	Obtenção das funções substitutas estimadoras de λ	62
3.1.1	Teste de validação	66
3.2	Propositura de heurísticas para a busca de redes com menor requerimento de λ via funções objetivo substitutas	68
3.3	Aplicações das heurísticas em experimentos computacionais	76
3.3.1	Encontrando redes alternativas a redes reais	76
3.3.2	Comparação da confiabilidade das redes de entrada e de saída	84
3.3.3	Começando por uma topologia aleatória	86
3.4	Conclusão	88
4	Conclusões e Trabalhos Futuros	89
	Referências	93
	Apêndices	102
	APÊNDICE A Notações e Definições	103
	APÊNDICE B Valores de $I(\lambda; k)$ para todo invariante k , com todas as redes juntas	110
	APÊNDICE C Valores de $I(\lambda; k)$ para todo invariante k , com as redes separadas por n	115
	APÊNDICE D Gráficos dos invariantes das Tabelas 2 e 3	123
	APÊNDICE E Tamanho da amostra em modelos lineares generalizados	151
	APÊNDICE F Aplicação das heurísticas nas redes reais	153
	APÊNDICE G Redes reais e redes soluções fornecidas pela heurística VNS-KDS	157

ANEXO A Evaluating Community Detection Methods in a Controlled Experiment **160**

 A.1 Introduction 160

 A.2 Methodology 162

 A.3 Results and discussion 164

 A.3.1 Detection of rings as communities in hierarchical ring networks . . . 164

 A.3.2 Analysis of the modularity index against the detected communities 168

 A.4 Conclusion 168

Lista de Abreviaturas e Siglas **170**

Lista de Símbolos **171**

Lista de Tabelas **172**

Lista de Figuras **174**

Resumo

As redes ópticas desempenham um papel vital na atual sociedade da informação, e isso coloca o projeto dessas redes como uma questão central. Um projeto ruim de uma rede óptica pode acarretar em desperdício de recursos e mau desempenho da rede. Muitos parâmetros podem indicar as características de uma rede, e dentre eles tem-se o número mínimo de comprimentos de onda (λ) necessários para atender uma determinada demanda de tráfego, que é um fator de custo dominante no dimensionamento de rede, onde a sua otimização maximiza o espectro disponível na rede. Uma modelagem natural para redes ópticas é por meio de grafos, que possuem um número de nós (n) e arestas (m). O espaço de possíveis redes cresce exponencialmente com n , o que dificulta a busca de redes que minimizem λ , e que é agravado pelo fato do cálculo de λ ser *NP-hard*. Com a hipótese de que o valor de λ é influenciado pela topologia da rede, busca-se então encontrar invariantes topológicos de grafos, de tempo computacional polinomial, que sejam bem correlacionados com λ , para que se possa estimar λ de um modo mais rápido, em função desses invariantes. No presente trabalho, procede-se então com uma busca exploratória de invariantes topológicos de grafos, no melhor dos esforços. Tal base de invariantes levantada é ranqueada, numa seleção de atributos inédita em redes ópticas, via estimadores de informação mútua. Para isso é usada uma amostra com $2,2 \times 10^6$ redes aleatórias que imitam redes reais, em que o ranqueamento dos invariantes ocorre com todas as redes juntas e também separadas por n . Como resultado, destacam-se aqui, os invariantes derivados do *edge betweenness*, que estão entre os mais bem posicionados nos *rankings* obtidos, demonstrando sua boa representatividade para explicar λ . De posse então, dos invariantes mais significativos para explicar λ , em seguida procede-se com regressões apropriadas para estimar λ . Tal estimativa facilita o teste de λ em uma grande quantidade de grafos, e é considerada em heurísticas para a busca, em poucos minutos, de topologias que minimizem o requerimento de comprimentos de onda. A economia total entre as redes reais de entrada e as respectivas redes de saída variou de 23% a 59% e, além disso, as redes de saída demonstram uma maior confiabilidade comparativamente às redes reais de entrada.

Abstract

Optical networks play a vital role in the current information society, and this puts the design of such networks as a central issue. Poor design of an optical network can lead to wasted resources and poor network performance. Many parameters can indicate the characteristics of a network, and among them there is the minimum number of wavelengths (λ) required to meet a given traffic demand, which is a dominant cost factor in network designing, where its optimization maximizes the spectrum available on the network. A natural modeling for optical networks is by means of graphs, which have a number of nodes (n) and edges (m). The number of possible networks grows exponentially with n , which makes difficult to find networks that minimize λ , what is aggravated by the fact that the calculation of λ is a NP-Hard problem. With the hypothesis that the value of λ to be influenced by the network topology, it is sought to find topological invariants of graphs with polynomial computational time, that are well correlated with λ , and so that λ can be estimated more quickly, as a function of these invariants. In the present work, it is proceeded with an exploratory search of graphs topological invariants, in the best of efforts. Such raised base of invariants is ranked, in an unprecedented selection of attributes in optical networks, via mutual information estimators. For this, a sample with 2.2×10^6 random networks that mimic real networks is used, where the invariants ranking occurs with all networks together and also separated by n . As a result, stand out the invariants derived from *edge betweenness*, which are among the best positioned in the obtained rankings, demonstrating their good representativeness to explain λ . Then, from the most significant invariants to explain λ , it is proceeded with appropriate regressions to estimate λ . This estimation facilitates the λ test in a large number of graphs and is considered in heuristics to search, in a few minutes, for topologies that minimize the requirement for wavelengths. The total savings between the real input networks and their output networks varies from 23% to 59% and, in addition, output networks demonstrate greater reliability compared to real input networks.

1 Introdução

Nos dias de hoje, o valor gerado pelo mercado está cada vez mais relacionado à obtenção, detenção e disseminação de informações, do que propriamente à produção de bens de consumo físicos, ou ainda à posse de propriedades valiosas pelas empresas, como era - pode-se dizer - até pouco mais de uma década atrás. Esse fato se comprova pela ascensão exponencial e recente de empresas de busca de informações como o Google, ou de redes sociais como o Facebook, que são avaliadas na escala de centenas de bilhões de dólares, despontando entre as empresas mais valiosas do planeta. Nesse recente e valioso mercado, as redes ópticas apresentam um papel vital diante de uma nova sociedade, que se coloca cada vez mais como ávida consumidora da informação rápida.

1.1 Redes ópticas e o número de comprimentos de onda

As redes ópticas são atualmente a base da extensa rede de comunicação digital global (ilustradas na Figura 1) devido a muitos fatores, como a grande capacidade de tráfego, velocidade e alcance. Nessas redes, muitos canais independentes podem compartilhar a mesma fibra óptica, aumentando a taxa de dados dentro da mesma infraestrutura, ou seja, muitos sinais podem ser transportados usando uma mesma fibra óptica. Cada sinal usa um comprimento de onda o qual deve ser diferente para cada sinal transportado. A totalidade desses diferentes comprimentos de onda que passam através das fibras de uma rede é comumente uma variável de interesse na literatura.

O presente trabalho trata especificamente de redes ópticas *backbone*, que é o tipo de rede que transporta dados coletados de redes menores que se conectam com ela. Uma rede *backbone* se comporta como uma espinha dorsal para o transporte de sinais advindos das redes menores. Em outras palavras, é uma rede na qual as redes locais ou regionais se conectam para interconexão de longa distância. Na Figura 2a, observa-se uma ilustração de interligação da rede óptica *backbone* NSFNET com pontos de redes regionais menores, e na Figura 2b, apenas a topologia física da referida rede *backbone*. No decorrer do texto que segue, ao se falar de redes ópticas, está-se falando de redes ópticas *backbone*.



Figura 1 – Ilustração representativa da extensa rede de comunicação digital global. Fonte: <https://pngtree.com> (2018).

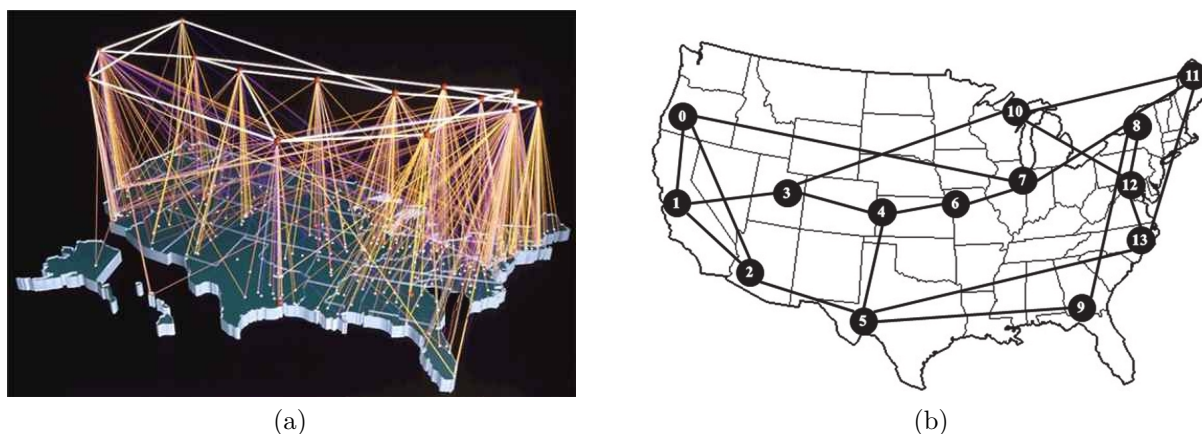


Figura 2 – Em (a), tem-se uma ilustração da interligação da rede óptica *backbone* NSFNET com pontos de redes regionais menores, e em (b) apenas a topologia física da referida rede *backbone*.

Nas redes ópticas convencionais, como as Redes Ópticas de Transporte (*Optical Transport Networks* - OTN, em inglês), a tecnologia de Multiplexação por Comprimentos de Onda (*Wavelength Division Multiplexing* - WDM, em inglês) permite que vários canais independentes compartilhem a mesma fibra óptica, o que proporciona um bom aproveitamento da banda de transmissão disponível nessas fibras, onde se torna possível lançar sinais com altas taxas de transmissão. Isso permite a implementação de Redes com Roteamento de Tráfego por Comprimentos de Onda (*Wavelength Routed Optical Networks* - WRON, em inglês) (BANERJEE; MUKHERJEE, 2000). Com a multiplexação de diferentes sinais forma-se o chamado agregado WDM, onde cada sinal tem seu comprimento de onda, diferentes entre si em cada aresta. Se conversores de comprimentos de onda não são considerados (hipótese adotada nesta tese), sinais reencaminhados opticamente mantêm seus comprimentos de onda (cores), restringindo a escolha das cores de sinais que compartilham arestas ao longo das rotas. Em uma rede em malha (*mesh network*) isso cria um problema complexo, largamente estudado na literatura (YANG; RAMAMURTHY, 2005; ZANG; JUE; MUKHERJEE, 2000), chamado de Roteamento e Alocação de Comprimentos de Onda (RWA - *Routing and Wavelength Assignment*).

O RWA pode ser resolvido à sua otimalidade, para redes relativamente pequenas, usando Programação Linear Inteira (*Integer Linear Programming* - ILP, em inglês), por exemplo como em Jaumard, Meyer e Thiongane (2007) e em Cousineau *et al.* (2015). Existem muitas abordagens com diferentes objetivos, mas a predominante é a que separa o RWA em dois subproblemas claros: roteamento da demanda de tráfego (definição dos caminhos de cada sinal), seguido de alocação de comprimentos de onda (definição dos comprimentos de onda de cada sinal), com o objetivo de minimizar o número de comprimentos de onda necessários para atender à demanda pelo roteamento atribuído. Essa função objetivo é normalmente escolhida no projeto de redes ópticas porque o requerimento de comprimentos de onda é relacionado com o custo e a capacidade das

redes. Fazendo um paralelo com a teoria de grafos, resolver o RWA buscando minimizar o número de comprimentos de onda, necessários para atender à demanda de tráfego, é equivalente a colorir o grafo de conflito, gerado pelo roteamento da demanda de tráfego, com a menor quantidade de cores possíveis.

O problema de RWA pode incluir a restrição de continuidade de comprimento de onda, o que o torna um problema *NP-hard* (ZANG; JUE; MUKHERJEE, 2000). Nesse contexto, cada canal óptico¹ deve usar o mesmo comprimento de onda desde o início até o final da rota. Como consequência, tal restrição pode gerar uma fragmentação espacial do espectro² disponível, que nada mais é do que a ocorrência de comprimentos de onda disponíveis em muitos enlaces, mas sem continuidade entre enlaces consecutivos, impedindo a criação de rotas com os comprimentos de onda excedentes que estejam isolados nos enlaces, como aparece ilustrado na Figura 3. Essa figura também ilustra a fragmentação espectral, a qual é tratada logo a frete.

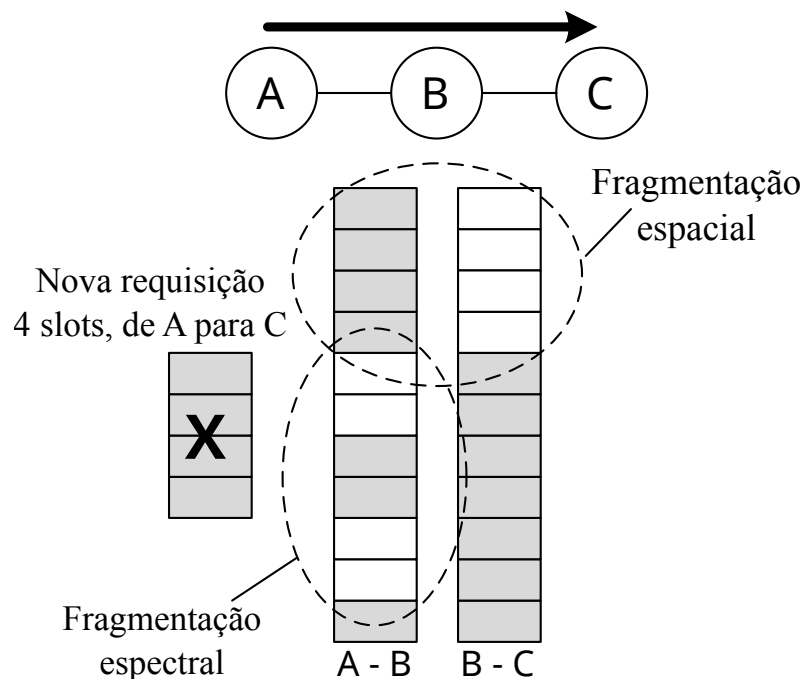


Figura 3 – Ilustração da fragmentação espacial e da fragmentação espectral. A barra da esquerda representa o espectro do enlace que liga A a B, e a barra da direita representa o espectro do enlace que liga B a C. Ao buscar atender uma demanda de tráfego de 4 *slots* de A até C, ocorre a fragmentação espacial, pois não há disponibilidade contínua de 4 *slots* em cada fibra, e ocorre ainda a fragmentação espectral, pois de A a B não há 4 *slots* contíguos (Figura adaptada de Izquierdo-Zaragoza, Pavon-Marino e Bueno-Delgado (2014) *apud* Tessinari (2016)).

¹A transporte de dados pelo meio óptico se dá pela transmissão de ondas eletromagnéticas, chamadas então de sinais ópticos. Um sinal óptico conectando dois pontos numa rede, sem conversão eletro-óptica ao longo do percurso, é chamado de canal óptico.

²Os dados, após serem processados eletronicamente, são convertidos em um sinal óptico com uma determinada frequência. A faixa de frequências utilizada para a transmissão nas redes ópticas é chamada de espectro óptico, ou simplesmente espectro.

O comprimento de onda é um recurso precioso em redes ópticas, portanto, uma característica importante de uma topologia é o **número mínimo de comprimentos de onda** necessários para estabelecer as demandas de tráfego (Marsan *et al.* (1993) *apud* Sheela (2012)). O termo **número mínimo de comprimentos de onda** é chamado de **número de comprimentos de onda** e denotado por λ . No presente estudo, λ é calculado de uma maneira exata usando o modelo de programação inteira, proposto por Cousineau *et al.* (2015) para resolver o RWA, como é citado na Seção 2.1.1. Na Figura 4, vê-se uma ilustração do conceito de número de comprimentos de onda, onde na Figura 4a três canais são transmitidos usando três comprimentos de onda, e na Figura 4b os mesmos três canais são transmitidos com apenas dois comprimentos de onda, que é a menor quantidade possível de comprimentos de onda que consegue atender a demanda de tráfego nesse caso, logo $\lambda = 2$.

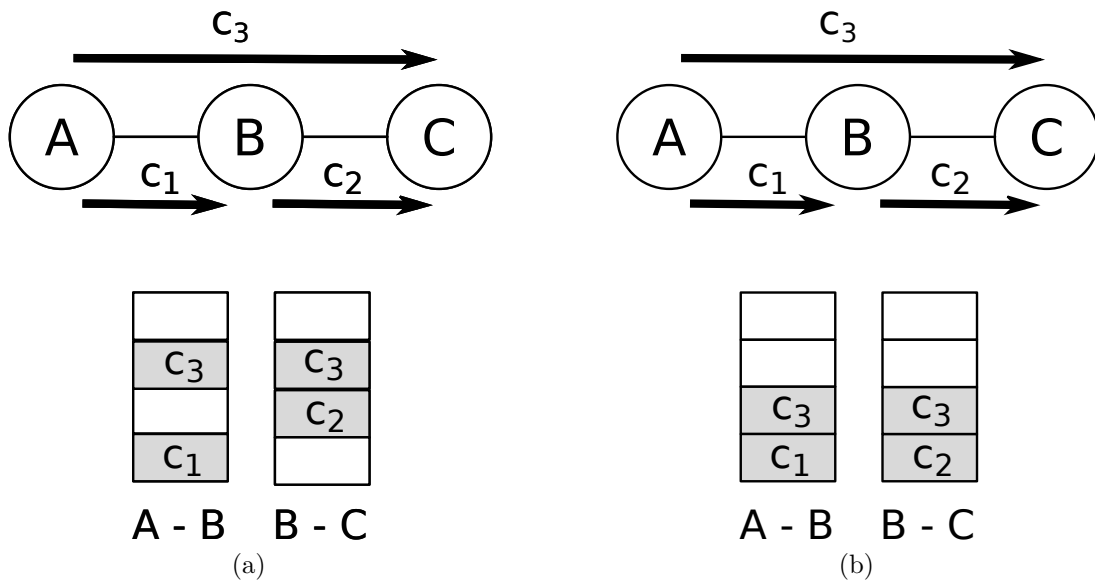


Figura 4 – Em (a) e em (b), três canais são transmitidos, onde em (a) essa alocação utiliza três comprimentos de onda, e já em (b) utiliza dois.

O número de comprimentos de onda usados para atender a demanda de tráfego é um fator de custo dominante no dimensionamento de rede, pois minimizar λ minimiza a fragmentação espacial, logo maximiza a disponibilidade espectral da rede, ou seja, maximiza a faixa de espectro não utilizada. Sendo assim, um projeto de topologia de rede eficiente, que permita diminuir λ , juntamente com as abordagens de roteamento, pode distribuir melhor o tráfego e evitar a sobrecarga de tráfego em enlaces específicos, reduzindo assim a possibilidade de *contention*³, conforme é sugerido por Aracil *et al.* (2007), *apud* Sheela (2012).

Uma nova geração de redes ópticas, chamadas de Redes Ópticas Elásticas (*Elastic Optical Networks* - EON, em inglês), baseada em Multiplexação por Divisão Ortogonal de

³Quando dois ou mais sinais usam o mesmo comprimento de onda.

Frequência Óptica (*Optical Orthogonal Frequency Division Multiplexing* - OOFDM, em inglês), permite um uso mais flexível do espectro óptico, com canais de diferentes tamanhos, alcançando maior eficiência espectral (ZHANG *et al.*, 2013; CHRISTODOULOPOULOS; TOMKOS; VARVARIGOS, 2011; TESSINARI *et al.*, 2016).

Em redes EON, existe o correspondente problema de Roteamento e Alocação do Espectro (*Routing and Spectrum Assignment* - RSA, em inglês), o qual trata da adição da restrição de contiguidade (*non-overlapping*, em inglês) ao problema RWA (PATEL *et al.*, 2012). Essa segunda restrição vem do fato de o espectro óptico ser, nesse caso, subdividido em *slots*, em que para atender a diferentes demandas de transmissão, canais de diferentes tamanhos são criados a partir de combinações dos *slots*, que devem permanecer contíguos (“colados”). A restrição de continuidade também se aplica aos canais, mas a combinação de canais com tamanhos diferentes causa um segundo tipo de fragmentação, a fragmentação espectral, dado que apenas os *slots* contíguos podem formar canais, o que gera “buracos” de frequência de espectro nos enlaces da rede, como pode ser visualizado na Figura 3. Uma estratégia comum para resolver este problema é a subdivisão do espectro em partições, atribuindo a cada uma apenas canais do mesmo tamanho. Assim, dentro de cada partição, o problema é reduzido ao RWA clássico (WANG; MUKHERJEE, 2014). De todo modo, minimizar λ em uma rede WDM é correlacionado a minimizar o número de canais usados em uma rede EON, como definido em Velasco *et al.* (2012). Tem-se ainda que, pelo fato do RSA ser um problema mais complexo, uma topologia com menor requerimento de comprimentos de onda facilitaria uma rede EON atender à restrição de continuidade, que é parte do problema RSA (TALEBI *et al.*, 2014).

1.2 O projeto de redes ópticas no contexto do presente trabalho

Independentemente do tipo de rede, um sistema inteligente para projetá-la muitas vezes envolve aspectos conflitantes como os recursos que a rede dispõe e os fluxos que ela deve atender, além ainda dos aspectos computacionais. Essa abordagem pode levar à otimização de um conjunto de atributos simultaneamente como em Yang *et al.* (2010), ou de diferentes tipos de fluxo de rede como em Przewoźniczek *et al.* (2015). Esses problemas são extensivamente estudados na literatura usando-se toda sorte de métodos de inteligência artificial, como redes neurais e algoritmos genéticos (HANAY; ARAKAWA; MURATA, 2015), dentre outros. Nesse processo de projeto de redes, modelos de programação linear também são frequentemente usados, como em Antunes, Craveirinha e Climaco (1993) e em Yoon, Baek e Tcha (1998).

A decisão de quais variáveis incluir no modelo do projeto de uma rede óptica é tão importante quanto, ou ainda mais importante que, o próprio modelo. Então, a escolha dos atributos a serem otimizados define o perfil da rede óptica obtida. Por exemplo, o

seu desempenho quanto a confiabilidade (*reliability*) (PAVAN *et al.*, 2015), a resiliência (*resilience*) (FRISKNEY; DONG; PHILLIPS, 2004), a probabilidade de bloqueio (*blocking probability*) (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015), as despesas de capital (*Capital Expenditure* - CAPEX) (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015), e ainda requerimentos de recursos, como o número de comprimentos de onda, dependem das variáveis escolhidas para serem otimizadas no processo de projeto da rede.

No presente trabalho, há o interesse particular na investigação de quais são os parâmetros topológicos das redes ópticas que melhor explicam o requerimento de comprimentos de onda. Não considera-se aqui, porém, a questão das distâncias físicas entre os nós, entraves geográficos, ou ainda os pontos levantados no paragrafo anterior, de modo que nos limitamos a explorar o requerimento de comprimentos de onda, visando maximizar a disponibilidade espectral da rede.

Uma técnica adequada para obter um conjunto de parâmetros que melhor explicam um determinado parâmetro escolhido é a seleção de atributos (*filter feature selection*, em inglês), um tópico amplamente explorado em muitas áreas de sistemas inteligentes (BENNASAR; HICKS; SETCHI, 2015). É realizado aqui um levantamento abrangente da literatura no que se refere a parâmetros topológicos de redes que são mais fáceis de calcular do que o número de comprimentos de onda. Então, é aplicada uma seleção de atributos, baseada num estimador eficiente de informação mútua, para classificar e selecionar os parâmetros que podem levar a um baixo requerimento de comprimentos de onda da rede (DEPIZZOL *et al.*, 2018). É utilizada uma estimação de informação mútua, ao invés de uma correlação linear simples, pois a informação mútua detecta o quão forte é a relação entre duas variáveis, independentemente da forma que se dá essa relação, ao invés da correlação linear que, como o próprio nome diz, só detecta correlações lineares.

Bennasar, Hicks e Setchi (2015) também trabalham com seleção de atributos baseada em informação mútua, mas no caso do presente trabalho, o estimador de informação mútua pode ser aplicado a dados discretos ou contínuos, independentemente, sem discretização ou qualquer conhecimento *a priori* a respeito da distribuição de probabilidade dos dados dos atributos tratados. Métodos de seleção de atributos têm vantagens como eficiência computacional e escalabilidade em termos da dimensionalidade do conjunto de dados. Já as desvantagens esperadas nesses métodos são a falta de informações sobre como se dá a interação entre os parâmetros (os atributos) e o classificador, além da ocorrência da seleção de parâmetros redundantes (BENNASAR; HICKS; SETCHI, 2015).

No presente trabalho, um estimador de entropia baseado na detecção de coincidências (MONTALVÃO; ATTUX; SILVA, 2014), é aplicado para definir o estimador de informação mútua usado no presente estudo. Este estimador é computacionalmente leve e adequado para espaços de alta dimensão, similar ao estimador *Neighborhood Mutual Information* (HU *et al.*, 2011). Essa nova abordagem usa o chamado “Método da Coincidência”, uma noção

emprestada da Mecânica Estatística (MA, 1985). Na Seção 2.2 tal estimador de entropia é apresentado para variáveis discretas, e também estendido para o caso contínuo. Como já foi exposto, a estimativa de informação mútua resultante permite medir a relevância entre atributos discretos e contínuos, sem discretização nem qualquer conhecimento *a priori* sobre a distribuição de probabilidade dos dados dos atributos analisados. Hu *et al.* (2011) também estimaram a entropia para atributos discretos e contínuos sem discretização, entretanto, a abordagem aqui apresentada, baseada no método de coincidência, é mais intuitiva e não precisa estimar funções de densidade de probabilidade.

Os parâmetros topológicos são então ranqueados com base em seus valores de informação mútua, obtendo um conjunto dos mais influentes para explicar o requerimento de comprimentos de onda. Como a informação mútua não pressupõe linearidade entre as variáveis (BENNASAR; HICKS; SETCHI, 2015), então os parâmetros relevantes são selecionados independentemente de como ocorre sua relação com o requerimento de comprimentos de onda. Os resultados desta análise servem como base para novos sistemas inteligentes de projeto e expansão de redes ópticas.

Alguns autores têm estudado parâmetros topológicos em redes WDM, como pode ser visto na Seção 1.4, mas com uma menor abrangência comparando-se ao presente trabalho, e sem qualquer resultado de classificação de parâmetros. No melhor dos esforços, não encontrou-se na literatura estudos relacionando atributos topológicos com qualquer outra característica de redes EON. No presente estudo, redes ópticas são modeladas por grafos, assim tal modelagem e alguns conceitos básicos necessários da teoria dos grafos são apresentados na Seção 1.3 a seguir.

1.3 Teoria de grafos para redes ópticas

Uma maneira intuitiva de representar redes ópticas é através de grafos. Um grafo é denotado por $G(V, E)$, ou apenas G , onde V é o conjunto de vértices e E é o conjunto de arestas (DIESTEL, 2016). Os vértices correspondem aos nós da rede e as arestas correspondem aos enlaces, de modo que as arestas conectam os vértices da mesma forma que os enlaces conectam os nós. O número de vértices é a ordem do grafo, denotado por $n = |V|$. O número de arestas é o tamanho do grafo e é denotado por $m = |E|$.

Dos grafos, é possível calcular os chamados invariantes de grafos, ou invariantes topológicos, que são parâmetros numéricos dos grafos que não mudam com a mudança dos rótulos de vértices ou arestas. Invariantes são importantes porque representam a topologia do grafo e, conseqüentemente, da rede óptica modelada pelo grafo. Assim, os invariantes de grafo são calculados neste estudo para entender como parâmetros topológicos da rede influenciam o número de comprimentos de onda (λ).

Um grafo $G(V,E)$ é conexo se, para cada par de vértices $u, v \in V$, existe pelo menos

um caminho de arestas conectando u e v . Um conjunto de corte de vértices de tamanho i de um grafo G , é um conjunto que contém i vértices de G que, quando retirados, desconectam o grafo. Analogamente, um conjunto de corte de arestas de tamanho i de um grafo G , é um conjunto que contém i arestas de G que, quando retiradas, desconectam o grafo. A conectividade de vértices (K) é definida como o tamanho mínimo de um corte de vértices de um grafo. Um grafo é chamado de 2-conexo quando, para cada par de vértices, há pelo menos dois caminhos disjuntos por vértices interconectando o par.

Diz-se que grafo 2-conexo tem conectividade de vértices igual a 2, e essa propriedade de 2-conexidade é conveniente para redes ópticas porque, no caso de uma única falha de vértice (ou de aresta), o grafo permanece conexo. Analogamente, tem-se a conectividade de arestas (K') que é definida como o tamanho mínimo de um corte de arestas. A conectividade de vértice de um grafo é menor ou igual a sua conectividade de aresta (DIESTEL, 2016). Um grafo conexo e sem ciclos é chamado de árvore. Uma árvore geradora de um grafo G , é um subgrafo de G que é uma árvore e que contém todos os vértices de G .

Grafos podem ter pesos em seus vértices ou arestas. Nesse caso, o grafo é chamado de ponderado, caso contrário, é chamado de não-ponderado. No caso de grafos não-ponderados, os caminhos mais curtos em termos de número de arestas são chamados de geodésicas, e a distância entre dois vértices $u, v \in V$ é definida como o comprimento de uma geodésica que conecta u e v .

A distância média de um grafo é a média das distâncias entre todos os pares de vértices do grafo, e a matriz de distância ($d_{i,j}$) tem com entrada, na linha i e coluna j , a distância entre os vértices i e j . Já a matriz de adjacência (A) possui entrada 1, na linha i e coluna j , se i é conectado diretamente com j com uma aresta, e zero caso contrário.

A distância de comunicabilidade (ESTRADA, 2012), C_d , é uma opção a distância euclidiana em grafos, onde $C_d[p, q] = G_{pp} + G_{qq} - 2G_{pq}$ com $G_{pq} = \sum_{j=1}^n \varphi_j(p)\varphi_j(q)e^{\lambda_j}$, onde $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ são os autovalores da matriz de adjacência A , e onde $\varphi_j(p)$ e $\varphi_j(q)$ são as p -ésima e q -ésima entradas do j -ésimo autovetor ortonormal de A associado com o autovalor λ_j . Por outro lado, propõe-se uma distância de comunicabilidade ajustada, AC_d , uma medida inspirada na distância de comunicabilidade (C_d), onde $AC_d[p, q] = G_{pp} + G_{qq} - 2G_{pq}$ mas com $G_{pq} = \exp(A_{pq})$. No presente trabalho é considerada AC_d , ao invés de C_d , uma vez que testes prévios demonstraram que AC_d é mais capaz de explicar o número de comprimentos de onda do que C_d .

Existem outras medidas do grafo que derivam da distância entre vértices, como a Transmissão. A Transmissão de um vértice i é dada pela soma de todas as distâncias de i para todos os outros vértices (CAPOROSI *et al.*, 2012). Outra medida é a Excentricidade, em que a Excentricidade de um vértice i é dada pela maior geodésica dentre as que ligam i a todos os outros vértices no grafo (SKIENA, 1990).

Além disso, dependendo da aplicação, as arestas podem ou não ter uma direção. Quando as arestas de um grafo têm uma direção, o grafo é dito direcionado. No contexto de redes ópticas, cada enlace permite fluxos em ambas as direções. Assim, apenas grafos não-direcionados são tratados neste trabalho.

Outra propriedade na teoria dos grafos é o grau do vértice. Para cada vértice i , o número de arestas conectadas a este vértice é chamado de grau do vértice i (DIESTEL, 2016). O grau médio é a média dos graus de todos os vértices de um grafo, e a variância de grau é a variância também dos graus de todos os vértices de um grafo. A matriz diagonal é uma matriz que assume o grau do vértice i na linha i e coluna i , e zero caso contrário.

A matriz laplaciana L é dada pela diferença entre a matriz diagonal e a matriz de adjacência. Do segundo menor autovalor da matriz laplaciana do grafo tem-se a conectividade algébrica (CHUNG, 1997). A matriz de distância de resistência (Ω_{ij}), para grafos não-direcionados, é dada por $\Omega_{ij} = (\Gamma)_{ii}^{-1} + (\Gamma)_{jj}^{-1} - 2(\Gamma)_{ij}^{-1}$, com Γ dado por $\Gamma = L + [1]_{(n \times n)}/n$ (LUKOVITS; NIKOLIĆ; TRINAJSTIĆ, 1999). O índice de Kirchhoff (PALACIOS, 2001) é definido pela metade da soma da matriz de distância de resistência (Ω_{ij}). Já o índice de Wiener (CAPOROSSI *et al.*, 2012) é definido pela metade da soma da matriz de distância ($d_{i,j}$). E o índice de Harary (PLAVŠIĆ *et al.*, 1993) é definido pela metade da soma do recíproco da matriz de distância ($1/d_{i,j}$).

Um particionamento do grafo é a sua divisão em partições de vértices, segundo algum critério prévio, onde todos os vértices do grafo estão em alguma partição, e não há interseção de vértices entre as partições. Se é buscado gerar as partições de modo que haja mais arestas dentro das partições do que entre elas ou, em outras palavras, que as ligações dentro das partições sejam mais fortes do que entre as partições, dizemos que essas partições formam comunidades. Dado um particionamento do grafo, é possível calcular o chamado índice de modularidade (CLAUSET; NEWMAN; MOORE, 2004), que informa o quão bom esse particionamento é, em termos da identificação de comunidades. O cálculo desse índice é dado na Equação A.1, do Anexo A, onde é explorado com mais detalhes.

Em um grafo, o *Edge betweenness* (EB_{ij}) é uma medida baseada na quantidade de menores caminhos que passam em uma aresta (Freeman (1979) *apud* Csardi e Nepusz (2006)). Assim, o *Edge betweenness* da aresta (i, j) é dada por $EB_{ij} = \sum g_{kl}^{ij}/g_{kl}; \forall k, l \in V, k < l$, onde g_{kl}^{ij} é o número de menores caminhos que ligam k e l , e que passam pela aresta (i, j) , e já g_{kl} é o número de todos os menores caminhos que ligam k e l . A Centralidade de *betweenness* ajustada de um vértice i é dada pelo soma do *edge betweenness* de todas as arestas ligadas ao vértice i , logo, mesmo sendo baseada em uma medida de arestas, a centralidade de *betweenness* ajustada é uma medida de vértices (CAPOROSSI *et al.*, 2012).

Analogamente, o *Vertex betweenness* (VB_i) é uma medida baseada na quantidade de caminhos que passam em um vértice (Freeman (1979) *apud* Csardi e Nepusz (2006)).

Assim, o *Vertex betweenness* do vértice i é dado por $VB_i = \sum g_{kl}^i / g_{kl}$; $\forall k, l \in V, k < l$, onde g_{kl}^i é o número de menores caminhos que ligam k e l , e que passam pelo vértice i , e já g_{kl} é o número de todos os menores caminhos que ligam k e l . A Centralidade de *vertex betweenness* é o *vertex betweenness* a nível de grafo, e é dado por $\sum_i [\max_j (VB_j) - VB_i]$; $i, j = 1, \dots, n$ (FREEMAN, 1979; WASSERMAN; FAUST, 1994).

Um grafo simples é aquele que (i) tem no máximo uma aresta para cada par de vértices, i.e., não tem arestas paralelas, e que (ii) não tem laços, i.e., não tem arestas que começam e terminam no mesmo vértice.

Seja G um grafo simples com n vértices e m arestas. O número máximo de arestas, m_{max} , em um grafo simples com n vértices é uma combinação de n tomados 2 a 2. Então, a densidade de arestas (α) de G é a razão entre m e m_{max} .

Na Seção 1.4 que segue, vê-se como alguns invariantes topológicos em redes WDM são tratados na literatura, via teoria de grafos, para explicar o número de comprimentos de onda.

1.4 Invariantes topológicos em redes WDM

Existem poucos trabalhos de pesquisa disponíveis na literatura que investigaram a relação entre invariantes topológicos e requerimento de comprimentos de onda em redes WDM (SHEELA, 2012). Que se tem notícia, Baroni e Bayvel (1997) foram os primeiros a analisar o requerimento de comprimentos de onda de redes reais e de um grande número de redes WDM conectadas aleatoriamente. As redes consideradas no referido estudo são simples, não-ponderadas, 2-conexas, e a densidade de arestas (α) satisfaz $0, 1 < \alpha < 0, 4$. A restrição de continuidade do comprimento de onda é assumida. Também é assumida uma demanda de tráfego uniforme, i.e., para cada par de nós $\{i, j\}$ da rede, há uma demanda única de i para j e uma de j para i , onde cada demanda usa uma geodésica, e um único comprimento de onda do começo ao fim da transmissão. Através de uma heurística, o número de comprimentos de onda das redes em estudo é estimado e depois analisado em relação a alguns invariantes, como número de vértices e densidade de arestas. Chama-se atenção para o fato de que nesse artigo a densidade de arestas é chamada de conectividade física. Os resultados do referido trabalho mostram que o requerimento médio de λ é praticamente independente da ordem n da rede, quando se analisa a densidade de arestas. Mais ainda, λ médio diminui fortemente à medida que a densidade de arestas aumenta.

Em Fenger *et al.* (2000), o requerimento de comprimentos de onda em relação aos parâmetros topológicos das redes WDM também é analisado. Para essa proposta, alguns milhões de redes aleatórias 2-conexas com um certo número de vértices e arestas são geradas, e os invariantes analisados dessas redes são o grau médio, a variância do grau e o

número de árvores geradoras. Resultados gerais são obtidos com valores médios de cada invariante topológico estudado. Nesse artigo, os autores também trabalham com grafos simples e não-ponderados, tráfego uniforme, roteamento usando geodésicas, e continuidade de comprimentos de onda. Também é considerado que não há limite no número de comprimentos de onda que podem passar por cada aresta. Os resultados obtidos para redes com 30 vértices e 45 arestas mostram que o λ médio aumenta com a variância do grau. Os autores observam que redes mais regulares, onde os vértices tendem a ter graus mais próximos, requerem menos comprimentos de onda. Por sua vez, observações em redes com o número de vértices e arestas iguais a 10 e 20, 20 e 30, 20 e 45, e 30 e 45, respectivamente, mostram uma relação inversamente proporcional entre o número de árvores geradoras e o λ médio. Tal relação também é observada para o grau médio, para a mesma faixa de ordem n . Uma diminuição considerável no requerimento de comprimentos de onda é encontrada com o aumento do grau médio para algo entre 4 e 5, sem vantagens significativas para redes com maiores graus médios. Portanto, usando a amostra analisada, o estudo concluiu que redes com grau médio entre 4 e 5 e com baixa variância de grau são os melhores possíveis em termos de requerimento de comprimentos de onda. Também é concluído que o número de árvores geradoras é um invariante muito preciso para medir a qualidade da rede em termos de eficiência de acomodação de tráfego, dado que as observações mostram uma relação de lei de potência entre o número de árvores geradoras e λ médio.

No estudo de [Châtelain et al. \(2009\)](#), dois conjuntos são analisados: um de 18 redes reais com número de nós variando de 11 a 53, e outro com $1,5 \times 10^6$ redes aleatórias com número de nós variando de 10 a 50, de 10 em 10. Esse estudo analisa apenas redes aleatórias com o número de arestas igual a 1,5, 2 e 2,5 vezes o número nós, o que implica numa faixa de grau médio entre 3 e 5. Para o conjunto de redes reais, é encontrado que $0,05 < \alpha < 0,45$, onde os valores mais baixos de α estão associados às maiores redes. O número de comprimentos de onda utilizados no estudo é estimado por um limite inferior (*lower bound*). Para redes aleatórias, uma relação de lei de potência é estabelecida entre o λ médio e a conectividade algébrica. Dessa relação, uma equação concisa é derivada para prever o requerimento de comprimentos de onda, a qual é testada em redes reais de longa distância. Os autores afirmam que a estimativa de requerimento de comprimentos de onda com base na conectividade algébrica é mais precisa do que as estimativas realizadas usando a variância do grau, o número de árvores geradoras e a distância média.

[Yuan e Xu \(2010\)](#) também estudaram o requerimento de comprimentos de onda, mas para redes ópticas com topologias físicas do tipo *small-world* e *scale-free* ([WANG; CHEN, 2003](#)), onde a distância média é analisada. O requerimento de comprimentos de onda também é estudado no caso da evolução da rede. Cerca de cem redes com 100 nós e 200 enlaces são analisadas. A demanda de tráfego usada para o RWA não é necessariamente *all-to-all*, os enlaces permitem fluxo bidirecional, e o roteamento não força o uso de geodésicas. As observações mostram que a distância média percorrida pelos sinais é diretamente

proporcional à distância média da rede. Ainda de acordo com as observações, redes com distâncias médias mais curtas tendem a exigir menos comprimentos de onda. Para essas redes, a taxa de crescimento do requerimento de comprimentos de onda é menor quando o volume de tráfego e a ordem da rede aumentam.

Enfatiza-se o fato de que, em quase todos os trabalhos discutidos anteriormente, (i) a média do número de comprimentos de onda é considerada, o que pode esconder casos extremos de requerimento excessivo ou reduzido de comprimentos de onda, e (ii) o número de comprimentos de onda é calculado usando heurísticas ou estimado por limites inferiores, o que pode levar a relações imprecisas entre os invariantes topológicos e o requerimento de comprimentos de onda. Em contraste, o presente trabalho considera, nas análises, redes com requerimento extremo (ou não) de comprimentos de onda, e não apenas a média dos requerimentos - média de λ dentre as redes, e os calcula como soluções otimizadas de um modelo ILP (COUSINEAU *et al.*, 2015) para todas as redes testadas. Estes são dois aspectos importantes da metodologia usada no presente trabalho, que o diferencia dos demais.

Na próxima seção, é apresentada uma delimitação do tema por meio de uma instância concreta do problema tratado no presente estudo via modelagem de grafos, o que induz à exploração de características topológicas de redes ópticas para estimar o requerimento de comprimentos de onda, e o uso dessa estimativa para gerar topologias com menores requerimentos de λ .

1.5 Delimitação do tema

O planejamento de uma nova rede óptica é um passo importante, uma vez que uma rede mal projetada pode desperdiçar recursos e operar com desempenho abaixo do potencial. Uma etapa crucial no processo de planejamento de rede é o projeto da topologia física da rede, que é quando se decide quais pares de nós serão conectados por enlaces de fibra óptica.

No projeto de redes ópticas, existe um grande número de topologias possíveis que podem ser geradas a partir de um número fixado de nós (n) e de enlaces (m), que são os parâmetros topológicos mais elementares. Por exemplo, existem 11.716.571 topologias conexas possíveis com 10 nós, das quais 9.743.542 são 2-conexas, valores estes calculados no programa *Nauty* (MCKAY, 2003). Por isso, não é fácil encontrar topologias que atendam a um conjunto de parâmetros desejados. Para ilustrar como o requerimento de comprimentos de onda depende dos parâmetros topológicos, a Figura 5 mostra duas topologias com o mesmo número de nós ($n = 10$) e enlaces ($m = 15$), mas com requerimento de comprimentos de onda diferentes para uma mesma demanda de tráfego uniforme⁴.

⁴Nesses dois casos da Figura 5, todas as considerações para o cálculo de λ são dadas na Seção 2.1.1.

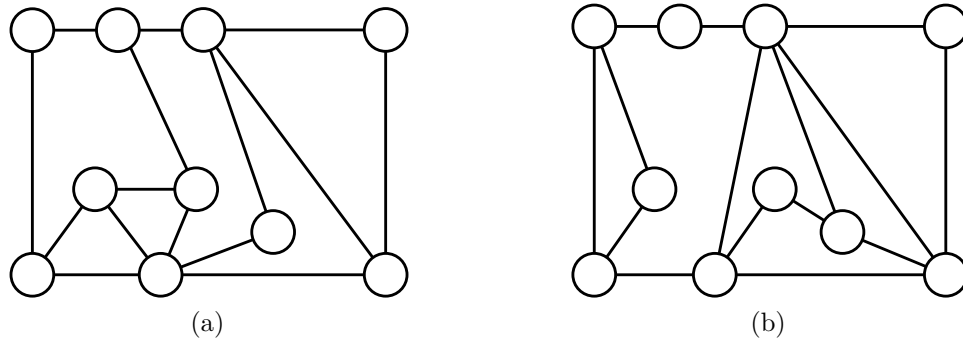


Figura 5 – Duas topologias de rede com o mesmo número de nós ($n = 10$) e enlaces ($m = 15$). A topologia (a) requer 8 comprimentos de onda, enquanto a topologia (b) requer 12 comprimentos de onda, i.e., 50% mais comprimentos de onda que a topologia (a).

Um projeto eficiente de redes ópticas é um desafio complexo e, dentre as várias maneiras de conectar nós usando enlaces, controlar requerimento de comprimentos de onda não é uma tarefa trivial. Para reduzir essa complexidade, em vez de usar diretamente os valores de requerimento de comprimentos de onda, cujo cálculo é um problema *NP-hard* (CHRISTODOULOPOULOS; TOMKOS; VARVARIGOS, 2011), é adequado buscar explorar os parâmetros topológicos das redes ópticas mais fáceis de serem calculados, e identificar como os valores desses parâmetros podem influenciar o requerimento de comprimentos de onda.

Assim, este estudo visa encontrar parâmetros topológicos, que podem ser obtidos com baixo custo computacional e que podem ser otimizados para levar a um baixo requerimento de comprimentos de onda. A informação mútua é usada para selecionar os parâmetros mais influentes nesse sentido. O conhecimento de tais parâmetros pode orientar melhor o projeto de topologias de rede na expansão de uma rede óptica atual ou na criação de novas redes. A lista completa dos parâmetros analisados no presente trabalho é dada no Apêndice A, o qual optou-se por se apresentar em língua inglesa buscando facilitar comparações e pesquisas futuras, fazendo com que essa lista contribua no sentido de ser um banco de consulta geral de parâmetros de grafos.

Num segundo momento, foca-se propriamente no projeto de topologia física de rede óptica (SIMMONS, 2014). O processo de projeto de topologias para redes ópticas é um problema complexo geralmente tratado por heurísticas, e a literatura explora essa abordagem de várias formas, tal como usando *simulated annealing* (XIAO; LEUNG; HUNG, 2001; GROVER; DOUCETTE, 2001; LIU; TOBAGI, 2008), algoritmos genéticos (BANERJEE; KUMAR, 2007; MORAIS *et al.*, 2011), e outros algoritmos evolucionários (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015; DUPLÉIX; ARAÚJO; BASTOS-FILHO, 2016). O projeto de topologias físicas através de abordagens heurísticas pode enfatizar apenas uma ou muitas funções objetivo diferentes, como o consumo de energia (DHARMAWEERA; PARTHIBAN; SEKERCIOGLU, 2011), custo (XIAO; LEUNG; HUNG, 2001;

CHAVES; BASTOS-FILHO; MARTINS-FILHO, 2010; ARAÚJO *et al.*, 2011; DHARMAWEERA; PARTHIBAN; SEKERCIOGLU, 2011), resiliência (DHARMAWEERA; PARTHIBAN; SEKERCIOGLU, 2011), probabilidade de bloqueio (CHAVES; BASTOS-FILHO; MARTINS-FILHO, 2010; ARAÚJO *et al.*, 2011), e Qualidade de Transmissão (QoT) (BASTOS-FILHO *et al.*, 2011; ARAÚJO *et al.*, 2011). Por exemplo, em Chaves, Bastos-Filho e Martins-Filho (2010) leva-se vários dias para concluir um projeto de uma rede óptica com 14 nós, em que se define a topologia física e se especificam os dispositivos ópticos que são implantados numa rede, buscando minimizar simultaneamente o custo total de capital e a probabilidade de bloqueio da rede. Em um avanço, no trabalho de Araújo *et al.* (2011) leva-se 24 horas para a mesma rede. Para evitar uma função objetivo custosa computacionalmente, em Araújo, Bastos-Filho e Martins-Filho (2015) usa-se funções objetivo substitutas que conciliam, de vários modos, questões conflitantes entre tempo e precisão. Em uma linha similar, o presente trabalho usa funções estimadoras como uma forma de mitigar problemas de maior requerimento de tempo computacional.

O projeto de topologias físicas de redes ópticas aqui proposto permite obter uma rede final com uma melhor alocação de caminhos ópticos, minimizando o requerimento de comprimentos de onda. Tais topologias permitem uma melhor alocação de caminho em redes WDM e em redes EON com um esquema de particionamento dedicado (WANG; MUKHERJEE, 2014; TESSINARI *et al.*, 2015) ou similar. Além disso, uma topologia que minimiza a fragmentação espacial pode ser de ajuda em outros cenários de EON. A fragmentação espacial é influenciada pela topologia física e é significativa mesmo para EON, quando a fragmentação espectral ainda é um problema (YIN *et al.*, 2013).

É proposto aqui então projetar topologias por meio de várias abordagens de heurísticas, onde cada solução visitada precisa ser avaliada por meio de uma função objetivo. Essa função objetivo também é aqui proposta, por meio de funções substitutas (HAN; ZHANG, 2012) obtidas a partir dos parâmetros topológicos de baixo custo computacional, encontrados como sendo os melhores correlacionados com λ , evitando assim o cálculo *NP-hard* exato de λ , nas muitas topologias que são visitadas a cada rodada da heurística.

A otimização baseada em funções substitutas é ainda uma técnica pouco explorada para projetar redes ópticas (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015). Os métodos via funções substitutas são construídos usando dados extraídos de cálculos dispendiosos obtidos por processos de alta fidelidade, e fornecem aproximações rápidas das funções objetivo em novos modelos, tornando assim os estudos de otimização mais viáveis (QUEIPO *et al.*, 2005).

Regressões estatísticas apropriadas são aqui usadas para criar tais funções substitutas, usando estimadores para valores esperados de λ ($E[\lambda]$) e valores mínimos esperados de λ ($E_{min}[\lambda]$). Os estimadores $E[\lambda]$ e $E_{min}[\lambda]$ são então as funções objetivo substitutas que, no presente trabalho, permitem criar heurísticas que encontrem topologias de redes ópticas

com um λ otimizado. Diferentemente de Araújo *et al.* (2011) e Chaves, Bastos-Filho e Martins-Filho (2010), a melhor das heurísticas aqui propostas encontra uma topologia melhorada com respeito a λ (para as ordens n testadas, entre 10 e 20) em, no máximo, poucos minutos.

Um resumo do que é tratado na presente seção, quanto a delimitação do tema desta tese, pode ser observado no diagrama dado na Figura 6.

Problema:

Projetar topologias físicas de redes ópticas, com requerimento de comprimento de onda (λ) minimizado



Dificuldade:

Calcular λ é um problema NP-hard

Alternativa proposta:

Buscar os parâmetros topológicos de redes ópticas que melhor expliquem λ ,
e
Estimar λ estatisticamente por meio desses invariantes topológicos



Usar essas estimativas para:

Buscar, por meio de heurísticas, topologias físicas de redes ópticas com menor requerimento de comprimentos de onda

Figura 6 – Diagrama que resume delimitação do tema da presente tese.

Ao surgir o interesse em estimar também outros parâmetros da rede que não o λ , mas também difíceis de calcular, outras funções objetivo alternativas para projetar topologias físicas podem ser obtidas usando a mesma abordagem usada para obter a função objetivo substituta que estima λ , com as devidas adaptações.

1.5.1 Objetivo Geral

Contribuir no projeto de topologias físicas para redes ópticas via uma abordagem estatística, de modo que tais redes tenham seu requerimento de comprimentos de onda minimizado, em função de parâmetros topológicos de tempo polinomial da rede óptica, a qual é modelada via grafos.

1.5.2 Objetivos Específicos

Diante do exposto, de modo geral, os objetivos específicos do presente trabalho podem ser listados como se segue:

- Fazer um levantamento da literatura, no melhor dos esforços, da maior quantidade possível de invariantes topológicos de grafos, e calculá-los para todos os grafos de uma amostra aleatória de $2,2 \times 10^6$ grafos que representam redes reais e também para uma amostra de redes reais;
- Aplicar o cálculo da informação mútua de λ junto a cada invariante de grafo levantado, gerando o ranqueamento de todos os invariantes considerados na análise, com respeito a quão bem eles são capazes de representar λ ;
- Encontrar funções substitutas de λ por meio de regressões apropriadas, utilizando-se dos invariantes de grafos mais bem posicionados no ranqueamento feito com respeito a λ ;
- Implementar diversas abordagens de heurísticas, utilizando-se de funções substitutas de λ como funções objetivo, para obter topologias com menor requerimento de comprimentos de onda.

1.6 Contribuições da tese

As contribuições desta tese tangem o projeto de topologia física de redes ópticas, modelando tais redes por meio de grafos, e aplicando análises estatísticas para buscar compreender como a topologia da rede pode influenciar o seu requerimento de comprimentos de onda. São propostas ainda heurísticas, baseadas em funções objetivo substitutas geradas nessas análises estatísticas, que retornam redes com menor requerimento de comprimentos de onda.

As topologias das redes são analisadas por meio dos invariantes topológicos, onde uma lista de 315 deles é considerada. Num primeiro momento contribui-se com o direcionamento de quais invariantes topológicos de rede são os mais relevantes quanto a serem capazes de explicar o requerimento de comprimentos de onda, segundo o ranqueamento considerado. Tal ranqueamento é dado por uma estimativa proposta da informação mútua, baseada no método da coincidência, o qual permite achar, de um modo menos complexo, uma correlação entre variáveis, independentemente da forma que se dá essa correlação. Para essa análise, é considerada uma amostra com $2,2 \times 10^6$ grafos aleatórios que simulam redes reais. Foram feitas análises considerando todas as redes juntas e ainda separadas pela ordem n , e achados foram contrastados com o que é estabelecido na literatura. Por fim, dos 315 invariantes iniciais, 20 deles se sobressaem, com destaque para o *Edge betweenness* máximo, que demonstrou a melhor correspondência com λ .

Num segundo momento, são propostas algumas abordagens de heurísticas para obter redes com demanda otimizada de número de comprimentos de onda, por meio de funções substitutas para λ , obtidas via regressões apropriadas com alguns invariantes topológicos que melhor explicam o requerimento de comprimentos de onda. Foram consideradas duas abordagens para a obtenção de topologias de redes com menor requerimento de comprimentos de onda. Na primeira abordagem, é gerada uma rede a partir de uma rede aleatória, e na segunda, a rede de entrada é uma rede real. Nesse segundo caso, foi significativa a diferença entre o valor de λ da rede de entrada e da de saída, chegando a reduções de até 59%. Foram comparadas ainda as confiabilidades das redes de entrada e de saída, onde as redes de saída, fornecidas pela heurística, demonstraram maior confiabilidade que as redes reais de entrada.

Por fim, vê-se que a metodologia aqui usada pode ser replicada para obter redes com outros parâmetros otimizados, também difíceis de calcular como o λ .

Em termos de publicações, todas as ligadas ao doutoramento, seguem listadas abaixo:

-> **Resumo em Congresso/Evento**

1) Silva, K.F.; Paiva, M.H.M.; Depizzol, D.B.; Segatto, M.E.V., “**A Study on Graph Invariants Applied to the Routing and Wavelength Assignment in Optical Networks**”, *International School and Conference on Complex Networks - NetSci-X*, Rio de Janeiro, 2015.

-> **Resumo Expandido em Congresso/Evento**

2) Depizzol, D.B.; Paiva, M.H.M.; Segatto, M.E.V., “**Analysis of the Relationship Between Topological Characteristics of Optical Networks and their Number of Wavelengths**”, *II Spring School on Networks*, Santiago, Chile, 2016.

-> **Artigo Completo em Congresso/Evento**

3) Silva, K.F.; Paiva, M.H.M.; Segatto, M.E.V.; Depizzol, D.B., “**Wavelength Requirements in Optical Backbone Networks Modelled as Twin Graphs**”, XLVIII SBPO - *Simpósio Brasileiro de Pesquisa Operacional*, Vitória, ES, 2016.

4) Depizzol, D.B.; Paiva, M.H.M.; Segatto, M.E.V., “**Evaluating community detection methods in a controlled experiment**”, *CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, DOI: 10.1109/CHILECON.2017.8229749, Pucon, Chile, 2017.

-> **Artigo completo em Revista Internacional [Qualis A1 em Engenharias IV]**

5) Depizzol, D.B.; Montalvão, J.; Lima, F.O.; Paiva, M.H.M.; Segatto, M.E.V., “**Feature Selection for Optical Network Design via a New Mutual Information Estimator**”, *Expert Systems With Applications*, DOI: 10.1016/j.eswa.2018.04.018, Elsevier, 2018.

Os trabalhos de 1 a 3 demonstram a exploração das primeiras relações buscadas entre invariantes de grafos e o requerimento de comprimentos de onda. Com essa base, um artigo completo foi escrito, e submetido à uma revista internacional, o que gerou a publicação 5.

O trabalho 4 surgiu da suspeita de que a presença forte de comunidades, medida pelo índice de modularidade, poderia afetar o requerimento de comprimentos de onda, porém testes com métodos de detecção de comunidades consagrados na literatura não chegavam a informações conclusivas, sequer sobre a presença de comunidades. Esse foi exatamente o ponto tratado nesse trabalho 4, onde foram confrontados os resultados de vários métodos de detecção de comunidades em redes hierárquicas, e muitas inconsistências foram detectadas. Vendo que esse questionamento inicial precisaria ser vencido, para então confirmar (ou não) a hipótese de interferência da presença de comunidades no valor de λ , percebeu-se que essa linha tomaria maiores proporções do que o planejado, e preferiu-se aguardar para abordá-la futuramente num momento mais oportuno. Como tal publicação foge da linha do que foi de fato abordado na presente tese, ela consta então, na íntegra, no Anexo A.

Um último artigo completo, que contempla a parte da aplicação de heurística via funções substitutas para obtenção de topologias com o requerimento de comprimentos de onda otimizado, está em processo de finalização, e também será destinado à submissão em revista internacional.

1.7 Organização da tese

Os principais resultados são apresentados nos Capítulos 2 e 3, onde o Capítulo 2 se trata da maior contribuição desta tese, e o Capítulo 3 pode ser considerado uma aplicação direta do Capítulo 2. No Capítulo 2 é apresentada toda a parte referente ao estudo de invariantes, além do ranqueamento desses invariantes via um método de estimação de informação mútua, detectando assim os invariantes de tempo polinomial que melhor explicam o requerimento de comprimentos de onda.

No Capítulo 3 são propostas funções substitutas para estimar λ , baseadas nos melhores invariantes detectados no Capítulo 2. A partir desses estimadores, heurísticas são propostas com o objetivo de encontrar topologias físicas de redes com menor requerimento de comprimentos de onda. O Capítulo 4 traz conclusões e propositura de trabalhos futuros.

2 Seleção de atributos para o ranqueamento de invariantes de redes ópticas

O presente capítulo, em resumo, trata de uma detecção inédita de invariantes topológicos de redes ópticas que melhor explicam o requerimento de comprimentos de onda. Tais invariantes são detectados via uma Seleção de Atributos, por meio de uma nova estimação da informação mútua baseada no método das coincidências, estimação essa que se mostra menos complexa que suas equivalentes. A motivação para essa detecção reside na possibilidade de usar esses melhores invariantes na busca de topologias físicas de redes ópticas que requeiram menos comprimentos de onda.

Em aprendizagem de máquinas e em estatística, Seleção de Atributos (“*Feature Selection*”, em Inglês) é um processo que resulta no ranqueamento e na seleção de um conjunto relevante de atributos, com respeito a uma variável de interesse. Também chamada de seleção de variáveis, ou de seleção de características, a seleção de atributos tem por objetivos principais: propiciar um melhor desempenho de previsão de variáveis de interesse, fornecer preditores mais rápidos e mais econômicos, e proporcionar uma melhor compreensão do processo subjacente que gerou os dados. Essa seleção de atributos, por consequência, gera uma redução da dimensionalidade espacial do problema considerado, o que o torna mais tratável (GUYON; ELISSEEFF, 2003; GUYON *et al.*, 2008).

Trabalhos de vanguarda sobre seleção de atributos, publicados no ano de 1997, usaram pouco mais que 40 atributos nos restritos domínios explorados (Blum e Langley (1997), Kohavi e John (1997) *apud* Guyon e Elisseeff (2003)). Em 2003, a situação já tinha tomado proporções exponenciais, onde muitos dos trabalhos já exploravam domínios com centenas a dezenas de milhares de atributos, como nos estudos crescentes dos genes que ganharam um fôlego considerável nessa época - por exemplo o mapeamento do genoma humano. Novas técnicas tem sido continuamente propostas para abordar essas tarefas desafiadoras envolvendo muitas variáveis possivelmente irrelevantes e redundantes e, muitas vezes, de posse de poucos exemplos de treinamento (GUYON; ELISSEEFF, 2003).

Os conceitos de redundância e irrelevância são distintos, uma vez que um atributo relevante pode ser redundante na presença de outro atributo relevante com o qual seja fortemente correlacionado (GUYON; ELISSEEFF, 2003). Por exemplo, em regressões lineares que envolvem dois ou mais atributos preditores, a manutenção de atributos redundantes ou irrelevantes acarreta no problema de aumento de variância nas respostas obtidas, e consequente instabilidade na modelagem (MONTGOMERY; PECK; VINING, 2012). James *et al.* (2013) inclusive afirmam que a seleção de atributos gera uma melhora considerável nas modelagens como um todo, propiciada justamente pela redução da

variância. Assim, a premissa central da seleção de atributos é que os dados podem conter características redundantes ou irrelevantes, que podem ser removidas sem perda considerável de informação e com ganho nas modelagens consideradas (BERMINGHAM *et al.*, 2015).

Existem muitos benefícios que podem ser obtidos por meio da seleção de atributos, dentre os quais podem-se citar: maior facilidade na visualização e na compreensão de dados, redução de requisitos de medição e de armazenamento, redução dos tempos de treinamento e utilização dos dados, e a já mencionada redução da dimensionalidade do problema. Essas vantagens são interessantes para a construção e seleção de subconjuntos de atributos que podem ser úteis para eleger ou construir bons preditores (GUYON; ELISSEEFF, 2003).

O método de seleção de atributos não pode ser confundido com o da Extração de Atributos (“*Feature Extraction*”, em Inglês). A extração de atributos (GUYON; ELISSEEFF, 2006) cria novos atributos em função dos originais, e a seleção de atributos (nosso caso) retorna um subconjunto com os atributos mais relevantes. Selecionar as variáveis mais relevantes geralmente é o ideal para construir um preditor, particularmente se as variáveis forem redundantes. Por outro lado, um subconjunto de variáveis úteis pode excluir muitas variáveis redundantes, mas relevantes (GUYON; ELISSEEFF, 2003).

Um método simples de ranqueamento de preditores pode ocorrer por meio do cálculo da bem conhecida correlação linear de cada um dos preditores com relação à variável de interesse, sobre a qual se deseja executar a predição. Esse método falha, porém, dado que algum preditor pode se relacionar fortemente com a variável de interesse, por meio de uma relação exponencial, por exemplo, e nesse caso a correlação linear padrão não detectaria devidamente tal relação, dado que ela apenas detecta correlações lineares, como o próprio nome diz. Assim, é desejável que a seleção de atributos executada não leve em consideração a relação entre as preditoras e a variável de interesse, detectando assim toda e qualquer relação significativa existente. O método usado para executar a seleção de atributos, apresentado na Seção 2.2, trabalha nesse sentido.

No presente caso de estudo de variáveis topológicas de redes ópticas, a variável de interesse aqui é o número de comprimentos de onda, que é uma variável valorizada no projeto de redes ópticas, uma vez que está relacionada aos seus custos. É tratado então um conjunto de 315 preditores, também referentes à topologia das redes ópticas, e busca-se saber, desse conjunto, quais são os que melhor explicam o número de comprimentos de onda em qualquer formato de relação. No melhor esforço que se pôde empreender, para efeito comparativo, não encontrou-se na literatura qualquer trabalho de ranqueamento de parâmetros topológicos de redes ópticas com respeito ao número mínimo de comprimentos de onda da rede. Encontrou-se apenas, trabalhos que tratavam da análise de alguns poucos invariantes topológicos de redes, os quais são apresentados na Seção 1.4.

Resumidamente, a metodologia adotada nesta etapa de seleção de atributos consiste

em gerar uma amostra aleatória de redes ópticas, que são representadas por grafos, e verificar nesta amostra quais invariantes topológicos têm maior influência no número de comprimentos de onda. Essa inspeção é realizada aplicando um estimador de informação mútua que é apresentado na Seção 2.2.

A literatura apresenta muitos métodos para calcular informações mútuas com um bom desempenho, como por exemplo o trabalho de [Bennasar, Hicks e Setchi \(2015\)](#). Optou-se pela estimativa da informação mútua dada pelo método da coincidência ([MA, 1981](#); [MA, 1985](#)), pois ele permite o cálculo da informação mútua de variáveis discretas e contínuas de forma simples, possibilitando a diferenciação dos invariantes mais importantes neste cenário, ao medir objetivamente a sua importância. Outra maneira de medir a informação mútua é através de funções de densidade de probabilidade (FDPs) das variáveis aleatórias em questão. No entanto, o método de coincidência é escolhido porque evita os problemas típicos da estimação de FDPs em espaços de alta dimensão, o que é o caso ao realizar associações de invariantes ([MONTALVÃO; ATTUX; SILVA, 2014](#)). O método da coincidência chama a atenção para a sua simplicidade, tanto na sua adaptabilidade a espaços de alta dimensão quanto na análise simultânea de variáveis discretas e contínuas combinadas (o que é nosso caso). Essa análise simultânea vem do fato de que é suficiente definir o evento “coincidência entre observações”, de modo que qualquer conjunto de variáveis (mesmo heterogêneas) possa ser analisado da mesma maneira e pelo mesmo método, como é explicado na Seção 2.2.

Tem-se então que, no presente capítulo, a Seção 2.1 descreve a amostra de redes reais e aleatórias consideradas neste estudo, bem como trata dos atributos a serem analisados. A Seção 2.2 detalha o método da coincidência usado para estimar a informação mútua, que é o ponto chave na seleção de atributos aqui considerada. Por fim, na Seção 2.3, são apresentados os resultados referentes a quais invariantes são mais interessantes para o projeto de topologias de redes ópticas, identificando os que melhor fornecem informações sobre o requerimento de comprimentos de onda.

2.1 As topologias de redes ópticas e os invariantes considerados para a seleção de atributos

Antes de proceder com a seleção de atributos em si, é necessário primeiramente obter valores dos atributos a serem trabalhados. Como os atributos aqui dizem respeito a medições em topologias de redes ópticas, tem-se então que obter amostras dessas topologias reais, ou que representem redes reais, para delas calcular os dados necessários. Como proposto, a modelagem de redes se dá aqui por meio de grafos, que são então nossos elementos de trabalho nessas amostras. Para tanto, trabalha-se com um conjunto de grafos aleatórios que simulam topologias reais de redes ópticas, os quais são explanados na Seção

2.1.1, e com um segundo conjunto de grafos que representam certas redes específicas do mundo real, mostrados na Seção 2.1.2.

2.1.1 As redes aleatórias e os invariantes considerados na análise

Esse primeiro conjunto é composto por grafos aleatórios que imitam redes ópticas reais, com número de vértices $n = 10, 11, \dots, 20$, onde 200.000 grafos são gerados aleatoriamente para cada n , no melhor dos esforços computacionais para gerar uma amostra tão grande quanto possível, resultando assim numa amostra com $2,2 \times 10^6$ grafos aleatórios. A opção pelo uso de grafos que simulam redes reais se dá pela própria limitação de obter uma amostra suficientemente grande de redes reais, para proceder com os cálculos que seguirão por todo este trabalho. Esses grafos aleatórios são gerados com ordem de pelo menos 10 nós, pois a partir desse valor de n a amostragem de grafos já passa a se tornar mais interessante, comparativamente a análise de todos os possíveis grafos, uma vez que a quantidade de todos os possíveis grafos para $n = 10$ já é grande o suficiente para tornar o trabalho menos factível computacionalmente. E é tomada ainda uma ordem máxima de 20 nós, como uma forma de limitar o tamanho da amostra, devido ao custo computacional envolvido. A taxa de elementos repetidos¹ na amostra é de 2%, ocorrido majoritariamente para $n = 10$. O programa usado para gerar tais grafos aleatórios foi o NetworkX (HAGBERG; SWART; CHULT, 2008), o qual era chamado via uma rotina em C (KERNIGHAN; RITCHIE, 2006), que otimizava o processo de geração dos grafos.

Além disso, para que esta amostra de grafos aleatórios seja o mais parecido possível com uma amostra de topologias de redes ópticas, são consideradas características comuns ou desejáveis às redes ópticas. Assim, os grafos aleatórios da amostra são gerados de modo que possuam as seguintes características:

- (i) simples, portanto, sem *loops* e sem arestas paralelas;
- (ii) não-orientado, ou seja, com arestas suportando fluxo bidirecional;
- (iii) com arestas de peso unitário, como forma de encontrar padrões que poderiam ser mascarados ao usar arestas com pesos não unitários;
- (iv) com arestas sem restrição de capacidade;
- (v) 2-conexos como em (BARONI; BAYVEL, 1997), apresentado na Seção 1.4. Dentre os grafos aleatórios gerados, apenas os 2-conexos são considerados, até obter 200.000 grafos por ordem de rede;
- (vi) com grau médio variando de 2 a 4, como esperado para as redes do mundo real (PAVAN *et al.*, 2010). O modelo de Erdős-Rényi (ERDÖS; RÉNYI, 1960) é usado para gerar

¹A amostra de $n = 10$ possui 19,8% dos elementos repetidos, a de $n = 11$ possui 2,4%, e a de $n = 12$ possui 0,3%. Para as demais ordens $n = 13, \dots, 20$ não houve repetição de elementos, e isso se justifica pela chance ser muito menor de escolha aleatória de um elemento repetido nestes casos de n maior, isso porque o espaço de possíveis redes a serem tomadas é muito maior.

os grafos aleatórios. Assumindo a hipótese de grau médio variando de 2 a 4, é fácil ver que a probabilidade de uma aresta conectar cada par de vértices em um grafo de ordem n é $3/(n - 1)$.

O método utilizado em todo o presente trabalho para resolver o RWA à otimalidade, e encontrar o valor de λ , é o dado por Cousineau *et al.* (2015), o qual combina modelos de programação linear inteira e limitantes inferiores (“*lower bounds*”, em Inglês), dentre outras técnicas, culminando numa abordagem computacional segmentada e de baixo custo que permite que o número de comprimentos de onda seja obtido mesmo para uma grande amostra.

O RWA é então resolvido para cada um dos $2,2 \times 10^6$ grafos da amostra aleatória, retornando o número de comprimentos de onda (λ) necessários para cada elemento da amostra. De todos os grafos dessa amostra, apenas 52 deles não conseguiram obter solução dentro do prazo estabelecido, que era de 5 segundos. Para resolver o RWA, é considerada uma demanda de tráfego uniforme, o que significa:

- (i) uma requisição de tráfego estática e bidirecional é imposta entre cada par de nós (demanda *all-to-all*);
- (ii) o roteamento é realizado exclusivamente usando caminhos mais curtos (geodésicas), considerando arestas com peso unitário; e
- (iii) a restrição de continuidade do comprimento de onda é imposta, o que significa que cada sinal deve ser transmitido usando um comprimento de onda fixo do nó de origem até o nó de destino.

Essas considerações para resolver o RWA são escolhidas para forçar um cenário extremo e fixo, de modo que com isso as conclusões obtidas possam ser extrapoladas para casos gerais. Uma vez que o interesse é entender como as topologias de rede podem interferir no requerimento de comprimentos de onda, é então necessário que todas as outras características permaneçam fixas, dado que as variações na demanda de tráfego ou no processo de roteamento, por exemplo, podem mascarar resultados.

A abordagem escolhida da análise topológica da amostra de grafos é dada por meio de medidas das topologias dos grafos, buscando determinar como os diferentes valores que essas medidas assumem podem influenciar no requerimento de comprimentos de onda. Algumas medidas topológicas referem-se a vértices ou arestas, e outras ainda ao grafo como um todo. Quando as medidas relativas ao grafo *per se* não dependem de rótulos de vértices ou arestas, elas são chamados de invariantes, como já foi mencionado. Caso contrário, elas são chamadas simplesmente de medidas. De modo resumido, pode-se dizer que o valor de um invariante varia em função dos grafos, mas é fixo quanto às rotulações de vértices e arestas desses grafos.

Um conjunto de medidas e invariantes descritos, tão diverso e tão grande quanto possível, é apresentado no Apêndice A, e muitos dos conceitos lá apresentados são usados no decorrer deste texto, sem apresentação prévia aqui. Algumas dessas medidas já são exaltadas na literatura de redes ópticas, e outras não foram ainda testadas na área. Uma abordagem exploratória é realizada, onde nenhum julgamento prévio é assumido sobre o potencial das medidas em explicar o número de comprimentos de onda. Essa abordagem é selecionada visando permitir que novas correlações tenham a oportunidade de serem identificadas.

Dado que no Apêndice A constam, além de invariantes, medidas que variam de acordo com a rotulação de vértices e arestas, para cada uma dessas medidas, são adotados os seguintes dez invariantes para cada grafo: (i) mínimo, (ii) máximo, (iii) média, (iv) mediana, (v) amplitude, (vi) distância interquartílica, (vii) desvio padrão, (viii) coeficiente de variação, (ix) desvio padrão mediano ou em relação à mediana² e (x) coeficiente de variação mediano ou em relação à mediana³.

Como pode ser observado no Apêndice A, existem 29 medidas e 30 invariantes, gerando com isso um total de $30 + (29 \times 10) = 320$ invariantes a serem analisados. As definições de diâmetro, distância máxima e excentricidade máxima coincidem, portanto apenas um deles (o diâmetro) é considerado. Também as definições de transmissão média e média de centralidade de *betweenness* ajustada coincidem, de modo que apenas uma delas (a transmissão média) é considerada. A distância mínima não é considerada porque é claramente sempre igual a 1. Portanto, a amplitude da distância também não é analisada, uma vez que a distância mínima é fixa e a distância máxima já é considerada. Com essas colocações, o número de invariantes analisados diminui para 315.

As medidas no Apêndice A são calculadas usando o pacote *igraph* (CSARDI; NEPUSZ, 2006) do programa R (R Core Team, 2016). R tem boa capacidade de lidar com grandes conjuntos de dados, e é um programa de código aberto, distribuído gratuitamente, e reconhecido pela comunidade científica, uma vez que são seus próprios membros que o desenvolvem e o atualizam continuamente. O pacote *igraph* contém dezenas de comandos já implementados para o cálculo de medidas de grafo e de invariantes.

Os invariantes topológicos calculados resultam numa matriz com 315 colunas (número de invariantes) por $2,2 \times 10^6 - 52$ linhas (número de grafos para os quais λ é calculado)⁴. Esta matriz não é apresentada aqui devido ao seu tamanho, mas é analisada posteriormente na Seção 2.3.

²Dado pela mediana das diferenças de todos os valores das medidas em relação à sua mediana.

³Dado pela razão entre valor obtido em (ix) e a mediana da medida, multiplicado por 100.

⁴O tempo total gasto no cálculo de todos os invariantes topológicos para todos os grafos foi de aproximadamente 331 horas, usando uma máquina com processador Intel Xeon E5-2430 v2 com 96 GB DDR3 de RAM.

2.1.2 As redes reais analisadas

Além das redes aleatórias dadas na seção anterior, um conjunto de 15 redes reais é considerado para análise em combinação com as aleatórias, com o objetivo de abordar as características de ambos os grupos. Este conjunto faz parte de um conjunto maior de 29 redes reais usadas em [Cousineau *et al.* \(2015\)](#), e essas 15 são selecionadas porque atendem às características da amostra aleatória (dadas na Seção 2.1.1), como o número de nós entre 10 e 20, e o grau médio entre 2 e 4, propiciando assim melhor efeito comparativo.

As redes reais também são representadas por grafos, e para cada uma, λ também é calculado à otimalidade usando o mesmo método aplicado às redes aleatórias. Todas as medidas do Apêndice A também são calculadas para as redes reais, resultando em uma matriz com 315 colunas (número de invariantes) e 15 linhas (número de redes reais), que não é apresentada aqui também devido ao seu tamanho. No entanto, ela é também analisada na Seção 2.3, juntamente com as medidas das redes aleatórias. A Tabela 1 mostra alguns parâmetros básicos dessas 15 redes reais selecionadas para análise.

Redes Reais	n	m	λ	Grau Médio
BREN	10	11	12	2,2
RNP	10	12	13	2,4
CESNET	12	19	15	3,2
VBNS	12	17	19	2,8
NSFNET	14	21	13	3,0
AUSTRIA	15	22	18	2,9
MZIMA	15	19	30	2,5
ARNES	17	20	38	2,4
GERMANY	17	26	24	3,1
SPAIN	17	28	22	3,3
CANARIE	19	26	44	2,7
EON	19	37	17	3,9
LAMBDA RAIL	19	23	58	2,4
MEMOREX	19	24	48	2,5
ARPANET	20	32	33	3,2

Tabela 1 – Alguns parâmetros básicos das 15 redes do mundo real analisadas.

2.2 Seleção de atributos baseada num novo estimador de informação mútua

Com a descrição das amostras das redes nas Seções 2.1.1 e 2.1.2 e com a delimitação das medidas no Apêndice A, nesta seção é apresentada a explicação do método usado para executar a seleção de atributos, ou seja, para escolher os melhores invariantes para o projeto de topologias para redes ópticas baseado na minimização do requerimento de comprimentos de onda. Este método foi elaborado numa parceria com o professor pesquisador Jugurta Montalvão da UFS, e está publicado em [Depizzol *et al.* \(2018\)](#), bem como os resultados dele advindos, no que se referem aos invariantes de grafos que mais se destacam no suporte

ao projeto de topologias de redes ópticas.

Para entender como esse método funciona, considera-se a análise do seguinte exemplo simples: um dado honesto possui C faces, com C desconhecido, e onde em cada face há uma letra diferente das demais, as quais não seguem ordenação alguma. Mesmo intuitivamente, se um observador tiver acesso a observações aleatórias e independentes do lançamento desse dado, tal observador pode tentar inferir C usando a enumeração de diferentes símbolos observados. Por exemplo, em 10 lances desse dado, observam-se as seguintes faces: f, u, p, o, w, j, t, t, l e p, nessa ordem. Com base nos símbolos observados, à primeira vista, parece que a melhor coisa que este observador intuitivo pode fazer é inferir que o dado tem pelo menos 8 faces.

Na verdade, o observador pode fazer muito mais, embora que de um modo menos intuitivo: observando que os símbolos t e p aparecem duas vezes, o observador pode contar a ocorrência de duas coincidências, uma para p e outra para t. Claramente, para identificar as duas coincidências, o observador deve, de alguma forma, ler a lista de letras e comparar cada uma das 10 com as outras 9. Desconsiderando as comparações repetidas, há um total de $(10 \times 9)/2 = 45$ comparações; ou seja, em 45 comparações, 2 coincidências são observadas. Analogamente, considerando agora um outro dado honesto com C faces, se apenas uma das faces é marcada como sendo “coincidência” e as demais $C - 1$ faces como “não coincidência”, espera-se uma ocorrência da face “coincidência” a cada C lances desse dado, em média. Portanto, C pode ser estimado como $\hat{C} = 45/2 = 22,5$ faces.

Em verdade, o experimento de lançamento de dados aqui considerado foi simulado computacionalmente considerando-se um dado honesto com $C = 20$ faces. Ressalta-se que a estimativa de aproximadamente 22 faces é baseada em apenas 10 jogadas de dados e, portanto, em um número de jogadas menores do que o número estimado de faces, o que é de certa forma contra-intuitivo.

Um fato notável é que, quando usado com dados desonestos, este método não estima o número real de faces, mas um número menor, que pode ser entendido como o número de faces efetivas ou o número de faces que um dado honesto deve ter para causar a mesma taxa de coincidência.

Embora esta apresentação seja muito simplificada, este princípio é a essência do método da coincidência proposto e usado por [Ma \(1981\)](#) no contexto da Mecânica Estatística, onde as coincidências são definidas em termos de trajetória de partículas em movimento, como explicado por [Ma \(1985\)](#), no Capítulo 25.

Ainda na analogia com dados, o número efetivo de faces, denotado por C , pode ser entendido como o número de estados equiprováveis de um sistema, que é chamado de “cardinalidade efetiva”. Dada uma amostra de saídas aleatórias e independentes, como as saídas do lançamento do dado honesto, o método da coincidência diz que C pode ser

estimado por uma “Cardinalidade Estimada” (\hat{C}), sendo definida por:

$$\hat{C} = \frac{\text{número de comparações}}{\text{número de coincidências}}.$$

Em termos de teoria da informação, a entropia quadrática (ou de colisão), denotada como H , associada à fonte aleatória (ou, no caso do lançamento dos dados, a medida da incerteza sobre qual letra é observada como resultado de um experimento de natureza aleatória) é simplesmente dado pelo logaritmo de C . Logo,

$$H = \log(C).$$

Analogamente, a entropia estimada é dada por

$$\hat{H} = \log(\hat{C}).$$

Neste exemplo com os dados, a observação {f, u, p, o, w, j, t, t, l, p} corresponde claramente a um fenômeno que é discreto e finito por natureza, porque as letras obtidas correspondem ao sorteio aleatório de algumas das C faces do dado. No caso de variáveis aleatórias contínuas, não há sentido em estimar a cardinalidade efetiva, pois ela sempre divergirá para o infinito, uma vez que usando o método da coincidência, entre as observações em escala contínua, a probabilidade de uma coincidência exata tende a zero. E com o número de coincidências tendendo para zero, ambos \hat{C} e \hat{H} tendem para o infinito.

No entanto, o método da coincidência ainda pode ser usado com uma redefinição adequada do significado de coincidência para medições contínuas. No caso, seria razoável assumir que duas medidas contínuas são coincidentes se a diferença absoluta entre elas for menor do que um determinado intervalo, chamado de vizinhança de coincidência, e denotado por Δ . Nesse caso, \hat{C} também dependeria do Δ selecionado na definição de coincidência. Quando Δ tende a zero, o número de coincidências também diminui, então \hat{C} tende para o infinito. Para evitar isso, \hat{C} é redefinido como cardinalidade de delta, \hat{C}_Δ , e dado por:

$$\hat{C}_\Delta = \hat{C} \times \Delta.$$

Quando Δ tende para zero, \hat{C} tende para o infinito, enquanto que \hat{C}_Δ pode convergir para um valor finito real se houver dados suficientes.

Analogamente, dado que \hat{H} também diverge para o infinito para Δ pequeno, uma entropia diferencial \hat{H}_Δ é definida por

$$\hat{H}_\Delta = \log(\hat{C}_\Delta) = \log(\hat{C} \times \Delta),$$

a qual pode ser reescrita como

$$\hat{H}_\Delta = \log(\hat{C}) + \log(\Delta).$$

No caso de observações multidimensionais, como no caso das associações de diferentes invariantes, a única adaptação necessária para o método de coincidência é uma definição adequada de Δ , que pode ser tão simples quanto um hipercubo, por exemplo. A seleção do volume de coincidência deve ser realizada com cautela, pois é o aspecto mais sensível do método de coincidência. O final desta seção trata de tal seleção.

Para explicar como a medida de H se aplica para estimar o número de comprimentos de onda, é conveniente retomar a analogia com a jogada de dados. Neste caso, dois dados são considerados: X com faces {a, b, c, d}, e Y com faces {e, f}. As cardinalidades efetivas de X e Y, C_X e C_Y , não são necessariamente 4 e 2, como inicialmente poderia ser esperado. Esta afirmação só é verdade se os dados forem honestos (faces equiprováveis). Porém, as cardinalidades efetivas C_X e C_Y podem ser estimadas eficientemente através do método da coincidência. No lançamento desses dois dados há $4 \times 2 = 8$ possíveis pares de observações resultantes. No entanto, a cardinalidade conjunta efetiva dos pares de observações, C_{XY} , não necessariamente é 8, pelo mesmo motivo. E ainda, C_{XY} também pode ser estimada usando o método da coincidência.

A influência mútua entre os dois dados pode ser medida por meio da cardinalidade de cada dado. Se os dados são independentes, C_{XY} deve ser igual ao produto das cardinalidades $C_X \times C_Y$. Por outro lado, a cardinalidade conjunta, C_{XY} , sempre é menor do que o produto das cardinalidades individuais C_X e C_Y no caso de dependência entre as variáveis (quando o resultado do dado X depende do resultado de Y, e vice-versa). Com isso é possível definir uma cardinalidade relativa (C_R) dada por

$$C_R = \frac{C_X \times C_Y}{C_{XY}}.$$

C_R pode assumir valores maiores ou iguais a 1, e $C_R = 1$ significa independência entre os dados. Por sua vez, quanto maior for o valor de C_R , maior será a relação de dependência entre X e Y. Na Teoria da Informação, $\log(C_R) = I(X; Y)$ é chamado de **informação mútua** entre as variáveis aleatórias X e Y (COVER; THOMAS, 2012), onde

$$\begin{aligned} I(X; Y) &= \log(C_R) \\ &= \log(C_X) + \log(C_Y) - \log(C_{XY}) \\ &= H(X) + H(Y) - H(X, Y). \end{aligned}$$

Com C_R assumindo valores maiores ou iguais a 1, conseqüentemente $I(X; Y) = \log(C_R)$ assume valores maiores ou iguais a zero.

Conforme explicado anteriormente, apenas a redefinição da coincidência é necessária no caso de variáveis contínuas, enquanto que todo o resto permanece o mesmo porque os intervalos, volumes e hiper-volumes usados na definição de coincidência se cancelam no cálculo de C_R .

A seleção de Δ realizada neste estudo está relacionada com a ideia proposta e justificada em [Montalvão, Attux e Silva \(2014\)](#), onde as vizinhanças de coincidência são definidas como uma função do desvio padrão (σ) das variáveis analisadas. Executando simulações empíricas com a Equação 12 do trabalho de [Montalvão, Attux e Silva \(2014\)](#), dada por $\hat{H}_\Delta - \hat{H}_{1,05\Delta} \approx 0,07$, obteve-se Δ que atende tal equação para o presente caso, tendo-se então que a vizinhança de coincidência do número de comprimentos de onda (Δ_λ) e do invariante k (Δ_k), com $k = 1, \dots, 315$, são então dadas por:

$$\Delta_\lambda = 0,2 \times \max\{\sigma(\lambda), 1\},$$

$$\Delta_k = 0,2 \times \max\{\sigma(k), 1\}.$$

Por sua vez, a vizinhança de coincidência para a cardinalidade conjunta efetiva dos pares de observações ($\Delta_{\lambda,k}$), que considera λ com cada invariante k , é dada por

$$\Delta_{\lambda,k} = 0,2 \times \max\{\sigma(k), \sigma(\lambda), 1\}.$$

Após esta introdução sobre o método da coincidência, é possível vê-lo em ação na Seção 2.3 a seguir, que trata dos resultados do cálculo da informação mútua $I(X, Y)$ com $Y = \lambda$ em função de invariantes topológicos $X = k$.

2.3 Resultados gerados a partir do ranqueamento de invariantes

Na presente seção são apresentados os resultados referentes à seleção de atributos topológicos de redes ópticas que melhor explicam o requerimento de comprimentos de onda. Tal seleção é obtida por meio de um ranqueamento desses atributos, executado pelo método da coincidência, o qual é apresentado na Seção 2.2. Na Seção 2.3.1, é mostrada a aplicação de tal método ao conjunto de redes aleatórias todas juntas, e na Seção 2.3.2 essa mesma análise é apresentada para as redes separadas pela ordem n , gerando ranqueamentos distintos em cada caso, os quais são descritos em detalhes. Na Seção 2.3.3, são analisadas relações entre alguns invariantes k e o comprimento de onda λ , além de $I(\lambda; k)$ em função da ordem da rede. Alguns resultados são contrastados com a literatura na Seção 2.3.4. E por fim, na Seção 2.3.5, são tratadas questões adicionais referentes ao projeto de topologia de redes ópticas, e uma compilação dos melhores invariantes é apresentada.

2.3.1 Análise de todas as redes de todas as ordens juntas

A informação mútua, $I(\lambda; k)$, é estimada para todos os invariantes $k = 1, \dots, 315$, com todas as redes aleatórias descritas na Seção 2.1.1, usando a metodologia proposta na Seção 2.2. Para tornar este cálculo viável para um conjunto tão grande de redes, toda a matriz de dados com 315 colunas (número de invariantes) por $2,2 \times 10^6 - 52$ linhas (número de redes) é dividida em 100 partes iguais (S_1, S_2, \dots, S_{100}) por meio de uma separação

sistemática. As linhas são numeradas de 1 a $2,2 \times 10^6 - 52$, onde a linha i se refere a invariantes da rede i (N_i). Portanto, as 100 partes são tomadas de tal maneira que:

$$\begin{aligned} S_1 &= \{N_1, N_{101}, N_{201}, \dots\}; \\ S_2 &= \{N_2, N_{102}, N_{202}, \dots\}; \\ &\vdots \\ S_{100} &= \{N_{100}, N_{200}, N_{300}, \dots\}. \end{aligned}$$

Então, $I(\lambda;k)$ é calculado para cada um dos 315 invariantes em cada uma das 100 partes da matriz. Procedendo dessa maneira é necessário um tempo computacional mais curto do que executar tal cálculo com todas as linhas de uma vez, ou seja, com todos os grafos de uma vez⁵. Em seguida, a média dos 100 valores de $I(\lambda;k)$, obtida para cada uma das 100 partes da amostra, é adotada para obter um único valor de informação mútua para cada invariante. A mediana, o máximo e o mínimo também são tomados. Uma tabela completa com tais resultados para todos os invariantes analisados pode ser vista no Apêndice B. Porém, na Figura 7, é possível observar o perfil gerado pela ordenação dos dados, e a faixa de valores de informação mútua obtidos pelos invariantes.

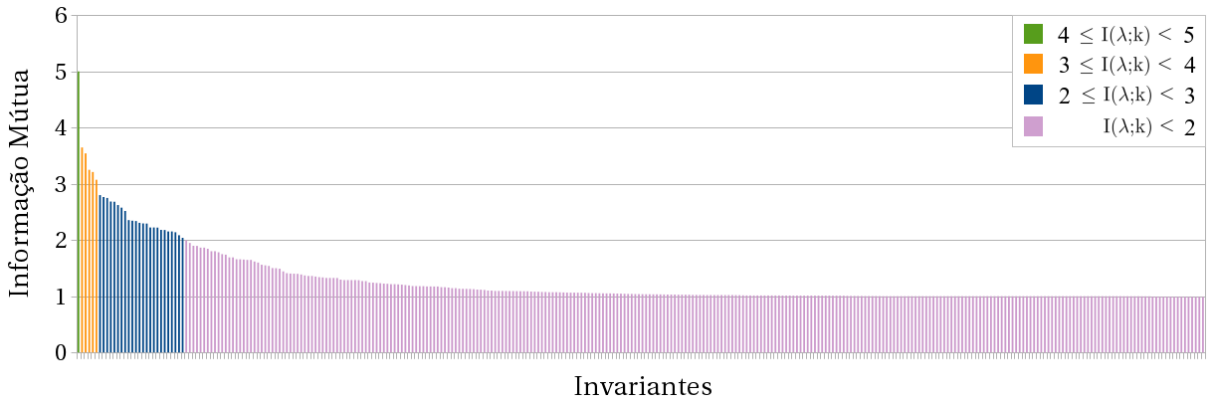


Figura 7 – Visualização do perfil gerado pela ordenação de todos os valores de informação mútua dos 315 invariantes analisados.

Se um invariante k e o número de comprimentos de onda λ tem informação mútua no intervalo, por exemplo, entre 1 e 2, i.e., $1 \leq I(\lambda;k) \leq 2$, então o conhecimento dos valores do invariante k podem reduzir a incerteza associada a λ em até a metade. Mais precisamente, a redução pela metade ocorre para $I(\lambda;k) = 2$. Dentre os 315 invariantes analisados, aqueles que reduzem a incerteza associada a λ em no mínimo a metade são os selecionados como relevantes, ou seja, os invariantes k com $I(\lambda;k) \geq 2$. Na Figura 7 observa-se que apenas uma pequena porção dos invariantes analisados atende a tal critério.

A Tabela 2 apresenta os únicos 30, de 315 invariantes, os quais obtiveram $I(\lambda;k) \geq 2$ para a média e para a mediana, listados em ordem decrescente com relação a esses valores, os quais também são apresentados na Figura 8.

⁵O tempo total gasto para o cálculo da informação mútua para todos os invariantes, somando os tempos gastos em cada amostra S_i , foi de aproximadamente 22 horas, no programa R; também usando uma máquina com processador Intel Xeon E5-2430 v2, com 96 GB DDR3 de RAM.

k	Invariantes	I(λ ;k)			
		Mínimo	Média	Mediana	Máximo
1	Edge betweenness máximo	4,92	4,9974	5,00	5,08
2	Desvio padrão do edge betweenness	3,57	3,6446	3,65	3,70
3	Amplitude do edge betweenness	3,47	3,5395	3,54	3,61
4	Índice de Kirchhoff	3,18	3,2476	3,25	3,32
5	Edge betweenness médio	3,15	3,2081	3,21	3,28
6	Vertex betweenness médio	3,02	3,0701	3,07	3,13
7	Distancia média	2,75	2,7981	2,80	2,84
8	Transmissão média	2,72	2,7639	2,76	2,81
9	Transmissão máxima	2,70	2,7470	2,75	2,80
10	Índice de Wiener	2,64	2,6837	2,68	2,73
11	Mediana da transmissão	2,63	2,6797	2,68	2,73
12	CV da distância de comunicabilidade ajustada	2,57	2,6202	2,62	2,67
13	Mediana do edge betweenness	2,53	2,5743	2,57	2,62
14	Índice de Harary	2,47	2,5154	2,52	2,58
15	Desvio padrão da transmissão	2,31	2,3529	2,35	2,40
16	Excentricidade média	2,30	2,3411	2,34	2,37
17	Amplitude da transmissão	2,30	2,3351	2,34	2,40
18	Coefficiente de variação da modularidade	2,26	2,3048	2,31	2,34
19	Máxima centralidade de betweenness ajustada	2,26	2,2940	2,29	2,34
20	Transmissão mínima	2,25	2,2909	2,29	2,33
21	Desvio padrão do vertex betweenness	2,19	2,2215	2,22	2,26
22	Desvio padrão da centralidade de betweenness ajustada	2,19	2,2215	2,22	2,26
23	Vertex betweenness máximo	2,19	2,2198	2,22	2,26
24	Centralidade de Vertex betweenness	2,15	2,1793	2,18	2,22
25	Distância de comunicabilidade ajustada média	2,15	2,1771	2,18	2,20
26	Amplitude do vertex betweenness	2,12	2,1508	2,15	2,19
27	Amplitude da centralidade de betweenness ajustada	2,12	2,1508	2,15	2,19
28	Número de vértices	2,10	2,1363	2,14	2,17
29	Distância interquartílica do edge betweenness	2,06	2,0841	2,08	2,11
30	Desvio padrão mediano do edge betweenness	2,01	2,0379	2,04	2,07

Tabela 2 – Os 30 invariantes que reduzem a pelo menos a metade a incerteza do número de comprimentos de onda, para as $2,2 \times 10^6 - 52$ redes aleatórias analisadas. Esses invariantes são ordenados em função de $I(\lambda;k)$ médio, e a cada um é atribuído um identificador “k” dado em função dessa ordenação. O Apêndice B exibe uma tabela com todos os 315 invariantes em estudo, ordenados por $I(\lambda;k)$ médio.

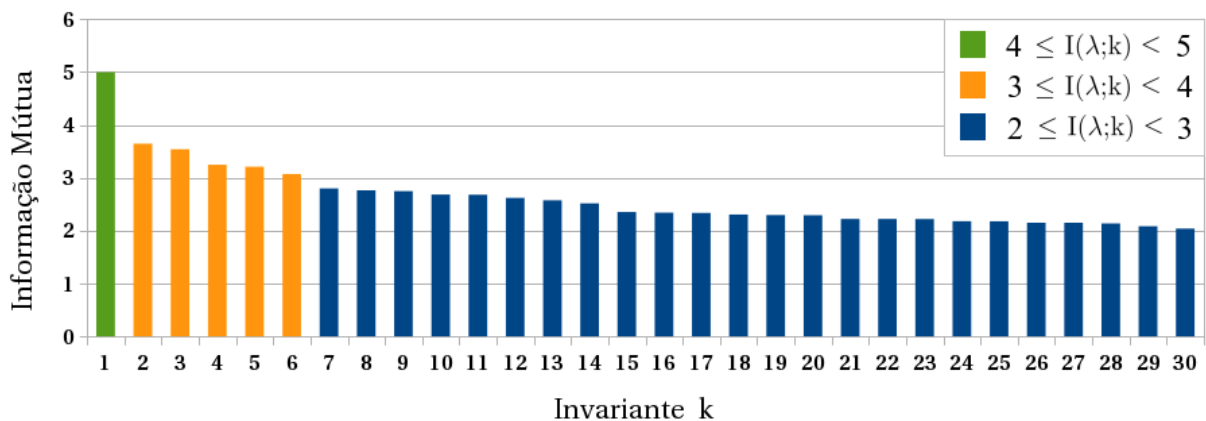


Figura 8 – Visualização do perfil gerado pela ordenação de todos os valores de informação mútua dos 30 invariantes que atenderam o critério de $I(\lambda;k) \geq 2$, ou seja, de reduzir a incerteza com relação à λ a, no mínimo, a metade.

Os invariantes que alcançaram a média de $I(\lambda;k)$ maior ou igual que 2, também o alcançaram para a mediana, e vice-versa. Ainda assim, de acordo com a Tabela 2, a proximidade dos valores mínimo, médio, mediano e máximo da informação mútua de cada invariante, dentre as 100 amostras, indicam que ela variou muito pouco de uma amostra pra outra, demonstrando a consistência dos resultados. Portanto, apenas a média das informações mútuas de cada invariante é considerada para o efeito de classificação, e o identificador k é atribuído aos invariantes de acordo com essa ordenação, como pode ser notado na Tabela 2, e igualmente no Apêndice B.

De acordo com a Tabela 2, o invariante que mais se destacou dentre todos é, de longe, o *edge betweenness* máximo ($k = 1$). A aresta que possui o valor máximo do *edge betweenness* é a aresta mais demandada pelos caminhos, sendo essa a aresta mais congestionada. *Edge betweenness* e *vertex betweenness* são chamadas aqui de medidas de congestionamento, e medem o quanto enlaces e nós são requeridos para atender a uma dada demanda de tráfego.

Os próximos dois invariantes na Tabela 2 são também relacionados com o congestionamento nas arestas, os quais são o desvio padrão ($k = 2$) e a amplitude ($k = 3$) do *edge betweenness*. Dado que os três primeiros invariantes que mais afetam o requerimento de comprimentos de onda são então baseados no *edge betweenness*, nos formatos de máximo, desvio padrão e amplitude, infere-se que é muito importante evitar a sobrecarga excessiva das arestas no projeto de redes ópticas. Outros invariantes da Tabela 2 que também refletem o congestionamento de arestas são os de $k = 5, 13, 29$ e 30 , o que confirma que o congestionamento de arestas é mesmo um fator importante neste estudo. O congestionamento por vértices também mostra ser um fator relevante, pois os invariantes $k = 6, 19, 21, 22, 23, 24, 26$, e 27 , que derivam do congestionamento de vértices, constam na Tabela 2. Como esses últimos invariantes citados ocupam posições mais baixas na tabela, isso sugere que o congestionamento por arestas é mais crítico do que o congestionamento por vértices, ao considerar o número de comprimentos de onda.

Notadamente, a transmissão também tem uma presença representativa, com os invariantes $k = 8, 9, 11, 15, 17$ e 20 na Tabela 2, o que destaca a importância dos vértices que são bons transmissores, i.e., aqueles cuja soma de todas as distâncias para os demais vértices é minimizada. Junto com a transmissão, a distância de comunicabilidade ajustada, representada pelo seu coeficiente de variação ($k = 12$) e sua média ($k = 25$), os índices de Kirchhoff ($k = 4$), de Wiener ($k = 10$), e de Harary ($k = 14$), a distância média ($k = 7$) e a excentricidade média ($k = 16$) enfatizam a dependência do número de comprimentos de onda com a distância entre vértices. Sobre a modularidade, ela está presente apenas com um invariante, seu coeficiente de variação ($k = 18$), o que indica que uma matriz de modularidade mais regular é desejável.

A Figura 9 ilustra os resultados a respeito do *edge betweenness* máximo. A Figura 9a apresenta o *edge betweenness* máximo versus o número de comprimentos de onda para ambos os conjuntos de redes aleatórias e reais em estudo, onde os losangos pretos representam os resultados das redes reais. A faixa colorida de hexágonos a direita é a legenda que reflete a faixa do quantitativo de grafos da amostra aleatória em cada hexágono do gráfico. Os mesmos resultados separados por número de nós são apresentados na Figura 9b, como forma de confirmar ou desmistificar os comportamentos observados na Figura 9a. Em ambas as figuras, ao invés de valores médios, todos os dados são apresentados, incluindo valores extremos do requerimento de comprimentos de onda encontrados. Destaca-se na Figura 9a que as redes reais apresentam um comportamento bem mais robusto e próximo do linear do que as redes aleatórias, eventualmente por causa do reduzido tamanho da amostra de redes reais, o qual diminui a chance de observar resultados mais extremos.

Na Figura 9c, é mostrada a informação mútua entre *edge betweenness* máximo e o número de comprimentos de onda para o conjunto de redes aleatórias em estudo, em função da ordem das redes, onde, nesse caso, observa-se que a importância do *edge betweenness* máximo para explicar λ decai com o aumento da ordem n rede. Para cada invariante na Tabela 2, esse trio de gráficos como o da Figura 9 é apresentado no Apêndice D.

Finalmente, é interessante notar a presença do invariante número de vértices ($k = 28$) na Tabela 2. O número de vértices versus o número de comprimentos de onda

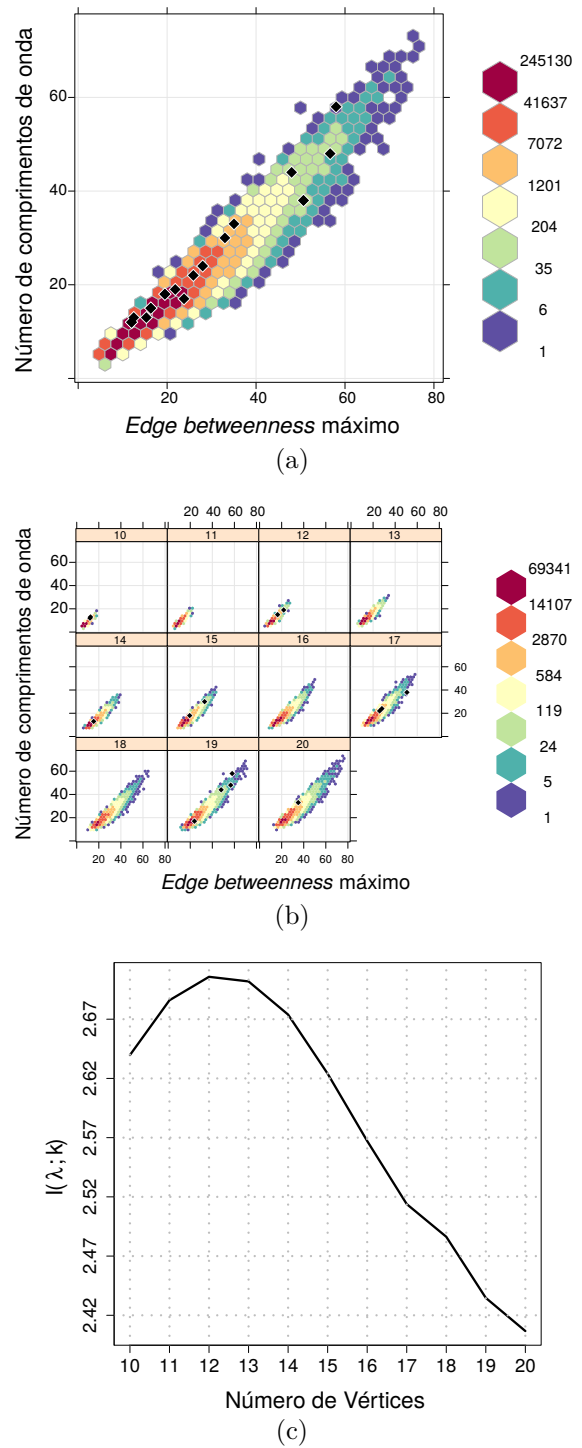


Figura 9 – *Edge betweenness* máximo ($k = 1$) versus λ para as redes reais (representadas por losangos pretos) e redes aleatórias em estudo: para todas as ordens de rede (a), e separadas por n (b). Em (c), informação mútua do *Edge betweenness* máximo e λ , em função da ordem da rede.

para todas as redes reais e aleatórias em estudo é mostrado na Figura 10, onde nesta figura observa-se que a ordem (n) é importante para ambos os conjuntos de redes, reais e aleatórios, onde redes maiores podem exigir mais comprimentos de onda do que redes menores, particularmente se eles tiverem graus médios mais baixos. Por outro lado, se o grau médio se mantiver alto, vê-se que entre redes menores e maiores a diferença no requerimento de comprimento de onda é menor, porém um grau médio alto em redes maiores se torna custoso dada a quantidade de arestas necessárias para manter alto o grau médio.

Em resumo, pode-se inferir que uma boa prática para o projeto de topologias de redes ópticas é dar tanta atenção quanto possível ao congestionamento, com o objetivo de minimizar extremos nos valores de *edge* e *vertex betweenness*, além de tornar os vértices da rede bons transmissores, ou seja, minimizar a soma das distâncias de cada vértice a todos os demais.

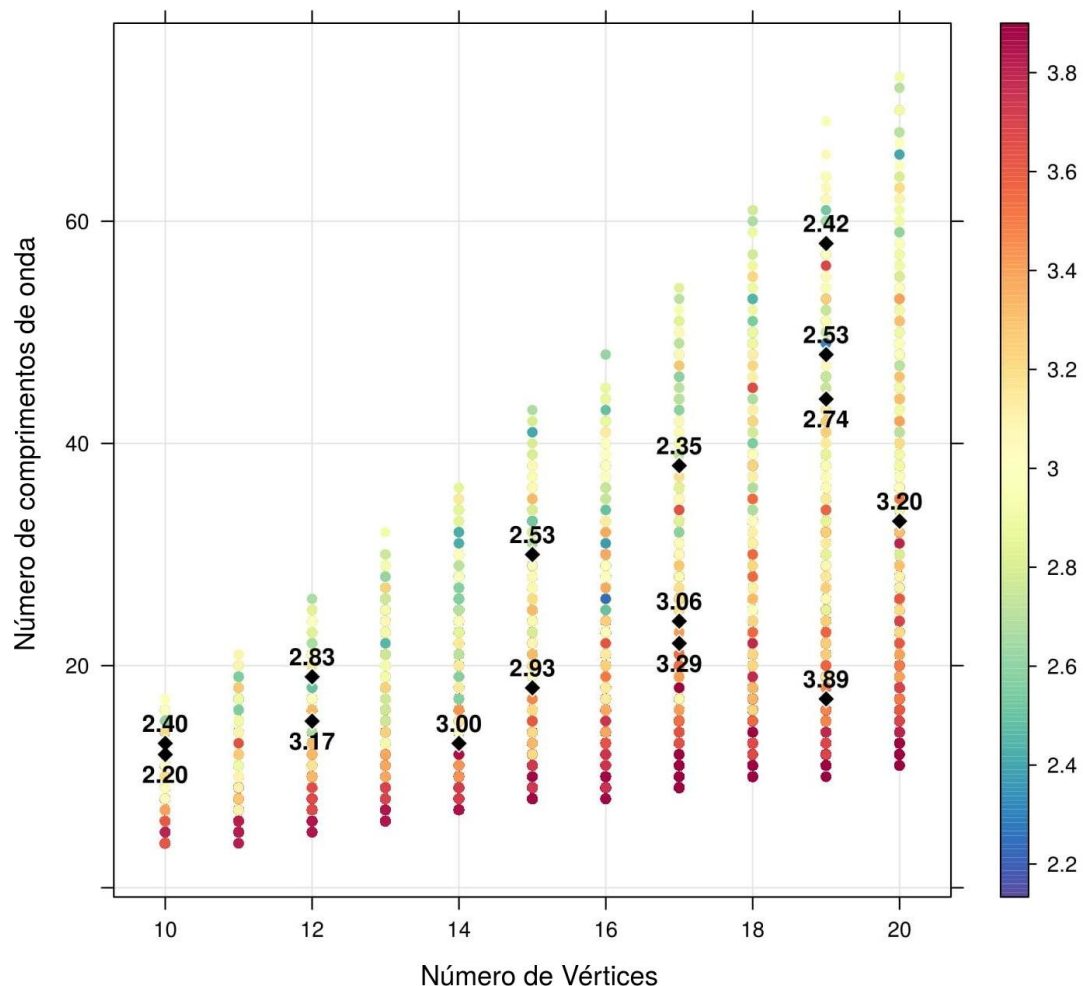


Figura 10 – O número de vértices ($k = 28$) versus λ , para todas as redes reais e aleatórias em estudo. As cores representam o grau médio, e o número em cada losango é o grau médio das redes do mundo real estudadas.

2.3.2 Análise das redes separadas pela ordem n

Nesta seção, a amostra total de grafos aleatórios é dividida em 11 amostras menores, em que os grafos na mesma amostra têm o mesmo número de vértices. O cálculo de $I(\lambda; k)$ é executado em cada uma dessas amostras para verificar (i) se, para cada n , a relevância dos invariantes pode ser diferente quando comparada à obtida com todos os grafos de todas as ordens juntos (visto na Seção 2.3.1) e (ii) se a relevância dos invariantes muda com o acréscimo ou o decréscimo da ordem n da rede. Semelhante ao que foi feito com a amostra inteira, cada amostra (das 11) é também dividida em 100 partes, e a média das informações mútuas de cada amostra é calculada para obter um valor de $I(\lambda; k)$ para cada invariante $k = 1, \dots, 315$.

O resultado deste processo pode ser visto na Tabela 3, o qual é apresentado para cada n independentemente. Como pode ser visto, poucos invariantes alcançaram $I(\lambda; k) \geq 2$, e portanto, as primeiras 30 posições foram tomadas para cada n , i.e., a mesma quantidade de invariantes tomada na Tabela 2, como uma forma de propiciar melhor comparação entre as duas análises. Uma tabela completa com a informação mútua de todas os 315 invariantes topológicos com as redes separadas por n é dada no Apêndice C. Os resultados em formato gráfico, para os invariantes apresentados na Tabela 3, são exibidos no Apêndice D, junto com os invariantes da Tabela 2.

Na Tabela 3, as células em cor branca indicam invariantes que constam, com maior ou menor valor de informação mútua, em todas as colunas, demonstrando assim maior importância na explicação de λ . Por outro lado, as células coloridas são aquelas cujos invariantes não constam em pelo menos uma coluna, e assim estes invariantes demonstram serem menos relevantes para explicar o λ . Alguns invariantes, com células coloridas ou em branco, não variam muito em altura dentre as colunas, não variando muito de importância para cada n , como, por exemplo, o invariante $k = 40$. Por sua vez, alguns invariantes mudam substancialmente suas posições em altura dentre as colunas (dentre as ordens de rede n), como o invariante $k = 13$, o qual aparece dentre os 30 mais importantes invariantes somente em redes menores.

De acordo com a Tabela 3, muitos invariantes que derivam do *edge betweenness* são também relevantes para cada n , especialmente o *edge betweenness* máximo, que segue sendo o mais importante inclusive em cada caso individual de n . A Figura 9b indica que a correlação positiva de *edge betweenness* máximo com λ observada na Figura 9a é mantida para cada n . Entretanto, na Figura 9c, observa-se que o valor da informação mútua $I(\lambda; k)$ decresce com o acréscimo de n .

A Tabela 3 mostra muitos invariantes que estão também presentes na Tabela 2: $k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 25, 29$, e 30. Portanto, 19 invariantes dos 30 da Tabela 2 seguem representativos com relação ao número de comprimentos de onda.

Tabela 3 – Os 30 invariantes mais importantes de acordo com os valores de $I(\lambda;k)$, para as $2,2 \times 10^6 - 52$ redes aleatórias analisadas, separadas pela ordem das redes. Células em branco indicam invariantes presentes em todas as colunas. Por outro lado, células coloridas indicam invariantes que não estão presentes em pelo menos uma coluna. O Apêndice C apresenta esta tabela de forma mais extensa, para o total dos 315 invariantes estudados.

n = 10		n = 11		n = 12		n = 13		n = 14		n = 15		n = 16		n = 17		n = 18		n = 19		n = 20	
k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)	k	I(λ;k)
1	2,6240	1	2,6860	1	2,7058	1	2,7019	1	2,6739	1	2,6240	1	2,5672	1	2,5139	1	2,4863	1	2,4347	1	2,4067
2	2,2566	40	2,1655	40	2,1882	2	2,2270	2	2,2543	2	2,2566	2	2,2636	2	2,2649	2	2,2676	2	2,2588	2	2,2500
40	2,2131	2	2,1082	2	2,1600	40	2,2221	40	2,2310	40	2,2131	40	2,1902	40	2,1707	40	2,1494	40	2,1276	40	2,1058
3	2,0277	4	1,9846	4	1,9947	3	2,0123	3	2,0216	3	2,0277	3	2,0269	3	2,0213	3	2,0200	3	2,0069	3	2,0032
4	1,9584	3	1,9226	3	1,9628	4	1,9976	4	1,9880	4	1,9584	4	1,9520	4	1,9431	4	1,9434	4	1,9088	4	1,9205
10	1,8414	10	1,8756	10	1,8879	10	1,8786	10	1,8770	10	1,8414	10	1,8261	10	1,8106	10	1,8190	10	1,7753	10	1,7796
8	1,8414	8	1,8756	8	1,8879	8	1,8786	8	1,8770	8	1,8414	8	1,8261	8	1,8106	8	1,8190	8	1,7753	8	1,7796
6	1,8414	6	1,8517	6	1,8712	6	1,8757	6	1,8770	6	1,8414	6	1,8261	6	1,8106	6	1,8190	6	1,7753	6	1,7796
14	1,7399	5	1,8372	14	1,8179	14	1,7929	14	1,7740	14	1,7399	14	1,7209	14	1,6959	14	1,7191	14	1,6697	9	1,6746
5	1,6754	14	1,7420	5	1,7921	5	1,7374	5	1,7263	5	1,6754	5	1,6704	9	1,6466	9	1,6593	9	1,6570	14	1,6677
9	1,6259	7	1,5971	7	1,6193	7	1,6323	9	1,6155	9	1,6259	9	1,6379	5	1,6438	5	1,6569	5	1,6122	5	1,6188
7	1,6139	59	1,5813	11	1,6112	9	1,6040	11	1,6115	7	1,6139	7	1,6144	7	1,6070	7	1,6087	16	1,5901	7	1,6015
16	1,5527	53	1,5497	59	1,5843	126	1,5738	7	1,6067	16	1,5527	11	1,5905	16	1,5840	16	1,6043	7	1,5859	57	1,5879
11	1,5516	52	1,5417	9	1,5630	59	1,5651	59	1,5357	11	1,5516	16	1,5699	11	1,5451	11	1,5990	11	1,5364	63	1,5862
59	1,5283	167	1,5417	52	1,5197	11	1,5446	16	1,5284	59	1,5283	101	1,5152	101	1,4972	101	1,5177	57	1,5010	61	1,5861
126	1,5279	12	1,5417	167	1,5197	16	1,4936	101	1,4973	126	1,5279	59	1,5134	59	1,4853	93	1,4881	101	1,5003	68	1,5859
101	1,4916	11	1,5405	12	1,5197	52	1,4903	52	1,4752	101	1,4916	93	1,4758	93	1,4775	59	1,4793	63	1,4990	16	1,5802
93	1,4548	9	1,5208	126	1,5085	167	1,4903	167	1,4752	93	1,4548	52	1,4441	126	1,4762	63	1,4674	61	1,4983	11	1,5745
52	1,4493	126	1,4955	101	1,4925	12	1,4903	12	1,4752	52	1,4493	167	1,4441	45	1,4421	61	1,4673	68	1,4982	101	1,5288
167	1,4493	101	1,4811	16	1,4685	101	1,4658	93	1,4398	167	1,4493	12	1,4441	115	1,4365	68	1,4660	93	1,4896	93	1,4977
12	1,4493	115	1,4700	25	1,4179	93	1,4331	45	1,4395	12	1,4493	45	1,4383	60	1,4301	57	1,4636	59	1,4595	59	1,4644
45	1,4408	16	1,4449	115	1,4104	45	1,4268	115	1,4136	45	1,4408	115	1,4331	52	1,4282	45	1,4499	126	1,4568	115	1,4535
115	1,4351	25	1,4236	45	1,4100	115	1,4093	126	1,4129	115	1,4351	29	1,3957	167	1,4282	115	1,4480	115	1,4458	45	1,4479
29	1,3898	109	1,4166	96	1,4077	29	1,3937	29	1,3960	29	1,3898	35	1,3834	12	1,4282	52	1,4321	45	1,4442	35	1,4162
34	1,3736	13	1,4114	13	1,3923	96	1,3845	96	1,3851	34	1,3736	34	1,3796	29	1,3987	167	1,4321	52	1,4076	52	1,4128
35	1,3736	46	1,4014	93	1,3854	13	1,3789	53	1,3760	35	1,3736	53	1,3755	35	1,3915	12	1,4321	167	1,4076	12	1,4128
96	1,3706	45	1,3871	109	1,3842	30	1,3704	30	1,3749	96	1,3706	30	1,3739	34	1,3834	35	1,4083	12	1,4076	15	1,4082
30	1,3671	96	1,3850	53	1,3783	138	1,3685	34	1,3656	30	1,3671	33	1,3722	33	1,3787	29	1,4068	35	1,4024	29	1,4049
33	1,3649	36	1,3719	46	1,3767	139	1,3685	35	1,3623	33	1,3649	96	1,3672	30	1,3749	60	1,4003	29	1,3994	167	1,4025
15	1,3390	138	1,3685	29	1,3687	34	1,3529	13	1,3607	15	1,3390	55	1,3645	63	1,3730	34	1,3998	60	1,3978	60	1,4011

Nome dos invariantes de acordo com k: 1: Edge betweenness máximo; 2: Desvio padrão do edge betweenness; 3: Amplitude do edge betweenness; 4: Índice de Kirchhoff; 5: Edge betweenness médio; 6: Vertex betweenness médio; 7: Distância média; 8: Transmissão média; 9: Transmissão máxima; 10: Índice de Wiener; 11: Transmissão Mediana; 12: Coeficiente de variação da distância de comunicabilidade ajustada; 13: Mediana do edge betweenness; 14: Índice de Harary; 15: Desvio padrão da transmissão; 16: Excentricidade Média; 25: Média da distância de comunicabilidade ajustada; 29: Distância interquartilica do edge betweenness; 30: Desvio padrão mediano do edge betweenness; 33: Coeficiente de variação do cocitation coupling; 34: Coeficiente de variação de dice similarity; 35: Coeficiente de variação de inverse log-weighted similarity; 36: Média de inverse log-weighted similarity; 40: Conectividade algébrica; 45: Desvio padrão da distância; 46: Média do cocitation coupling; 52: Número de arestas; 53: Centralização do grau; 55: Excentricidade Mediana; 57: Coeficiente de variação mediano de dice similarity; 59: Número de conjuntos mínimos que desconectam o grafo; 60: Coeficiente de variação da excentricidade; 61: Coeficiente de variação mediano de Jaccard similarity; 63: Coeficiente de variação mediano de inverse log-weighted similarity; 68: Coeficiente de variação mediano de cocitation coupling; 93: Coeficiente de variação de edge betweenness; 96: Mínimo de subgraph centrality; 101: Mínimo do grau médio dos vizinhos mais próximos; 109: Número de árvores geradoras; 115: Coeficiente de variação da distancia; 126: Coeficiente de variação mediano da transmissão; 138: Coeficiente de variação mediano de closeness centrality; 139: Coeficiente de variação mediano de closeness centrality normalizado; 167: Grau médio.

Esses 19 invariantes se localizam em posições superiores da Tabela 3 e são basicamente medidas de congestionamento e distância. Há ainda outros invariantes $k = 45, 55, 60, 93, 115,$ e $126,$ que também se remetem a questões de congestionamento e distância, e aparecem na Tabela 3, mas não na Tabela 2. Esses fatos confirmam que congestionamento e distância são tópicos que merecem atenção especial ao se projetar topologias de redes ópticas. Observa-se, no entanto, que dentre todos os invariantes de congestionamento na Tabela 3, apenas um se refere ao congestionamento por vértices ($k = 6$), enquanto que 5 invariantes dessa categoria ($k = 6, 21, 23, 24,$ e 26) constam na Tabela 2. Isso significa que o congestionamento de vértice não é tão relevante quanto o congestionamento de arestas no contexto do projeto de topologia de redes ópticas.

Dentre os 19 invariantes mencionados acima, 14 deles ($k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14,$ e 16) aparecem com maior importância para todo n , e as 5 restantes ($k = 13, 15, 25, 29,$ e 30) constam em uma ou mais (mas não todas) colunas na Tabela 3. Analisando as ordens das redes separadamente, esses 14 invariantes provam serem mais robustos, pois além de constarem em todas as colunas da Tabela 3, eles também mantêm aproximadamente o mesmo nível de importância para cada n , ou seja, mantêm aproximadamente a mesma altura dentre as colunas. Dentre os invariantes que não constam em todas as colunas, $k = 13,$ e 25 são importantes em redes menores e $k = 15, 29,$ e 30 não apresentam um padrão claro.

A Tabela 4 mostra a matriz de correlação linear para os 14 invariantes mencionados. Baseando-se nesta tabela, é possível observar que muitos desses invariantes são fortemente correlacionados, especialmente aqueles que derivam da mesma medida, como o par de invariantes $k = 8$ e $k = 11$, os quais derivam da medida de transmissão, ou o par $k = 1$ e $k = 3$, que derivam do *edge betweenness*.

Quando duas variáveis são ambas muito importantes (com alta informação mútua), derivadas da mesma variável, e fortemente correlacionadas, recomenda-se selecionar apenas uma delas, nesse caso, aquela com maior valor de informação mútua, porque isso evita

Invariantes	k	1	2	3	4	5	6	7	8	9	10	11	12	14	16
<i>Edge betweenness</i> máximo	1	1,00	0,96	0,98	0,91	0,91	0,91	0,91	0,88	0,89	0,85	0,88	-0,87	0,73	0,86
Desvio Padrão do <i>edge betweenness</i>	2		1,00	0,97	0,85	0,85	0,85	0,89	0,81	0,83	0,77	0,81	-0,80	0,63	0,86
Amplitude do <i>edge betweenness</i>	3			1,00	0,85	0,84	0,84	0,85	0,82	0,83	0,79	0,81	-0,79	0,67	0,82
Número de arestas	4				1,00	0,98	0,99	0,92	0,99	0,98	0,98	0,98	-0,97	0,89	0,84
<i>Edge betweenness</i> médio	5					1,00	0,98	0,97	0,96	0,94	0,93	0,96	-0,98	0,82	0,90
<i>Vertex betweenness</i> médio	6						1,00	0,94	0,99	0,98	0,98	0,99	-0,97	0,90	0,87
Distância média	7							1,00	0,90	0,90	0,85	0,90	-0,94	0,71	0,96
Transmissão média	8								1,00	0,99	0,99	1,00	-0,97	0,94	0,83
Transmissão máxima	9									1,00	0,98	0,98	-0,95	0,92	0,85
Índice de Wiener	10										1,00	0,99	-0,94	0,96	0,77
Transmissão Mediana	11											1,00	-0,97	0,94	0,82
CV da distância de comunicabilidade ajustada	12												1,00	-0,87	-0,85
Índice de Harary	14													1,00	0,62
Excentricidade média	16														1,00

Tabela 4 – Matriz de correlação linear dos 14 invariantes mais representativos, i.e., aqueles que aparecem na Tabela 2 e em todas as colunas da Tabela 3.

possíveis redundâncias e ajuda na simplificação da escolha dos invariantes. Por exemplo, no contexto de redes ópticas, não há necessidade de considerar ambos *edge betweenness* máximo ($k = 1$) e amplitude de *edge betweenness* ($k = 3$), porque eles têm correlação linear igual a 0,98.

Os demais invariantes que não aparecem na Tabela 2, mas constam na Tabela 3, são as de $k = 33, 34, 35, 36, 40, 46, 52, 53, 57, 59, 61, 63, 68, 96, 101, 109, 138, 139, 167$ num total de 19 invariantes adicionais. Dentre eles, 5 invariantes constam na Tabela 3 para todas as ordens de redes: conectividade algébrica ($k = 40$), número de arestas ($k = 52$), número de conjuntos mínimos que desconectam o grafo ($k = 59$), mínimo do grau médio dos vizinhos mais próximos ($k = 101$), e grau médio ($k = 167$). A conectividade algébrica se destaca, estando dentre as primeiras posições, enquanto os demais invariantes ocupam posições de meia altura na Tabela 3.

Dentre os 19 invariantes adicionais, três deles se remetem ao grau do vértice ($k = 53, 101, e 167$), com duas delas constando em todas as colunas de n ($k = 101, 167$), como dito anteriormente. Isso mostra a importância das medidas relacionadas ao grau para explicar o número de comprimentos de onda, como afirmado em Fenger *et al.* (2000). Observa-se, porém, que nenhum invariante de grau consta na Tabela 2.

É interessante destacar também a presença do invariante número de arestas ($k = 52$) na Tabela 3, para todas as ordens de redes, mostrando a importância intuitiva desse invariante, em que mais arestas implicam em um menor valor de λ (e.g., uma rede completa tem λ igual a 1). O invariante conectividade algébrica ($k = 40$), embora não apareça na Tabela 2, permanece entre as primeiras posições para todas as ordens de redes na Tabela 3, mostrando que a relação entre a conectividade algébrica e λ é realmente importante, como observado por Châtelain *et al.* (2009).

Por outro lado, o invariante número de árvores geradoras ($k = 109$) é aclamado por Fenger *et al.* (2000) mas não demonstrou grande importância uma vez que ele está presente em apenas duas colunas da Tabela 3 ($n = 11$ e 12) e em baixas posições. Mais ainda, esse invariante sequer consta na Tabela 2. Já o invariante número de conjuntos mínimos que desconectam o grafo - CMDG ($k = 59$) parece ser importante porque consta em todas as colunas da Tabela 3. O número de CMDG é relacionado com a conectividade da rede porque uma falha em um vértice de um CMDG influencia na chance da rede se tornar desconexa devido a outras falhas.

A dispersão de *closeness centrality* é retratada em dois invariantes da Tabela 3 ($k = 138, e 139$), mas não de forma muito representativa, dado que constam apenas para duas ordens de rede e em posições muito baixas nas colunas. O *cocitation coupling*, que reflete conexões de pares de vértices com um terceiro vértice, é representado por sua média no invariante $k = 46$ e por sua dispersão em $k = 33$ e em $k = 68$, que também têm posições baixas e aparecem para poucas ordens de rede.

A *similarity*, que retrata o quão são semelhantes os vértices (em aspectos definidos por cada tipo de similaridade), é apresentado na Tabela 3 para os invariantes $k = 34, 35, 36, 57, 61$ e 63 , todos medidas de dispersão, com exceção do 36 , que é uma média. Os invariantes $k = 57, 61$ e 63 aparentam ganhar importância em redes ópticas maiores. Por sua vez, os invariantes $k = 34$ e 35 constam em posições baixas, e o invariante $k = 36$ só aparece para $n = 11$.

Observações na Tabela 3 também indicam, porém, invariantes que crescem em importância com o aumento da ordem da rede, que são as medidas de dispersão de *log-weighted similarity* ($k = 35$ e $k = 63$), de *dice similarity* ($k = 57$), de *Jaccard similarity* ($k = 61$), e de *cocitation coupling* ($k = 68$). Porém, analisando o último gráfico de cada trio apresentado no Apêndice D, é possível constatar mais precisamente o ganho (ou não) de importância de cada invariante, das Tabelas 2 e 3, com o crescimento da ordem n da rede.

O *subgraph centrality* de um vértice mede o número de subgrafos em que um vértice participa. O invariante mínimo de *subgraph centrality* ($k = 96$) é mais representativo para redes menores, perdendo importância com o crescimento de n .

Finalmente, existem invariantes que aparecem na Tabela 2 mas não na Tabela 3. Claramente, o invariante número de vértices ($k = 28$) não aparece na Tabela 3, dado que ele é usado para separar a amostra de redes em colunas por n . Alguns invariantes baseados em amplitude ($k = 17, 26$ e 27) também se tornam menos importantes na Tabela 3, provavelmente porque as análises restritas a cada ordem de rede podem ter gerado amplitudes com menos variabilidade, e logo com menos sensibilidade para responder à variação de λ . O único invariante que envolve a modularidade ($k = 18$) aparece na Tabela 2 mas não na Tabela 3, mostrando a baixa importância deste invariante neste estudo. A Figura 11 mostra um resumo de toda a análise feita na presente seção, mostrando os tipos de invariantes que saem e entram em cena em cada caso.

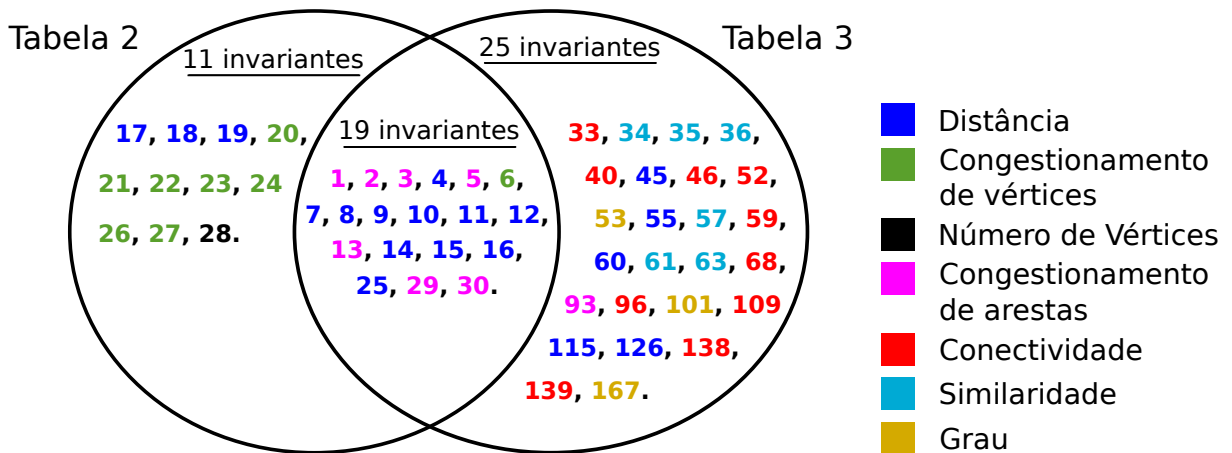


Figura 11 – Comparação dos invariantes que foram mais importantes na análise com todas as redes com ordens n juntas e separadas.

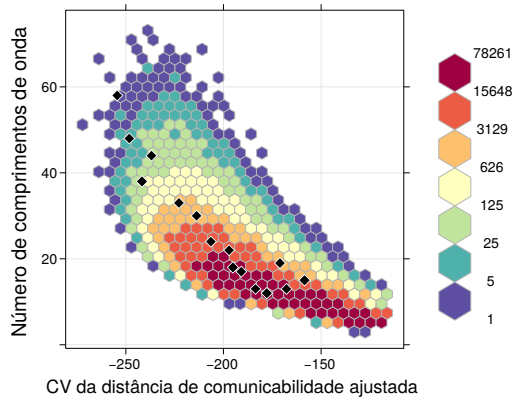
Em resumo, a Tabela 2 mostra, e a Tabela 3 confirma, a importância de considerar variáveis de congestionamento e de transmissão para o projeto de topologias de redes ópticas. Da análise da Tabela 3, observa-se que mais atenção deve ser dada ao congestionamento de arestas do que ao de vértices, e que além disso recomenda-se considerar: o grau médio em conjunto com o número de arestas, além dos invariantes de conectividade: conectividade algébrica e o número de conjuntos mínimos que desconectam o grafo; e ainda alguns fatores de similaridade ($k = 57, 61, \text{ e } 63$) em caso de expansão de rede (aumento da ordem n). Chama-se a atenção para as invariantes de conectividade, que têm predominância quando se faz a análise das redes separadas por n , uma vez que retirando-se de cena o fator ordem da rede n , abre-se caminho para o destaque desses invariantes.

2.3.3 Resposta de λ a alguns invariantes, e de $I(\lambda; k)$ a ordem n da rede

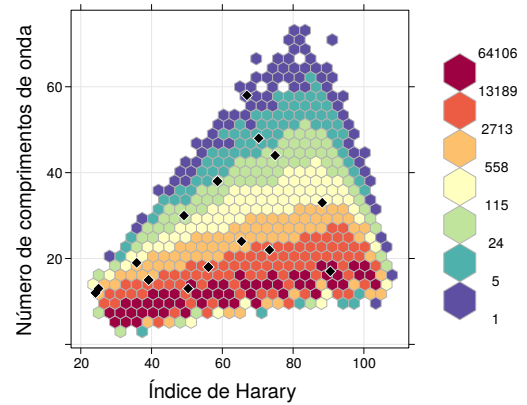
Ao analisar mais de perto o comportamento de alguns invariantes perante λ , se por um lado, o *edge betweenness* máximo ($k = 1$) tem uma correlação positiva com λ , tanto com todas as redes de todas as ordens n juntas (Figura 9a) quanto para redes com cada n separadamente (Figura 9b) - o que torna interessante que redes ópticas tenham baixos valores desse invariante; por outro lado, o coeficiente de variação da distância de comunicabilidade ajustada ($k = 12$), por exemplo, tende a ter uma correlação negativa com λ (Figuras 12a e 12b), de modo que altos valores do invariante $k = 12$ são interessantes para a tentativa de minimizar λ . Observa-se, no entanto, que a relação do invariante $k = 12$ com λ nas amostras de redes separadas por n (Figura 12b) apresentam padrões visuais diferentes entre si, e estes são diferentes ainda da relação dada na amostra de redes de todas as ordens juntas (Figura 12a). E, assim como para o *edge betweenness* máximo, a importância do coeficiente de variação da distância de comunicabilidade ajustada diminui com o crescimento da rede (Figuras 9c e 12c).

Um invariante que merece atenção é o Índice de Harary ($k = 14$). A Figura 13a pode sugerir uma correlação positiva do invariante $k = 14$ com λ , ainda que fraca. Entretanto, uma análise da Figura 13b indica o oposto, i.e., uma relação indireta para cada n . Esse comportamento contrário em cada caso demonstra a importância de considerar análises para cada n . A Figura 13c indica um decréscimo da importância do Índice de Harary, dado o crescimento da rede, quando se busca explicar λ com esse índice. É possível observar a tendência de correlação (positiva ou negativa) de cada invariante das Tabelas 2 e 3 com λ , pela análise dos gráficos no Apêndice D.

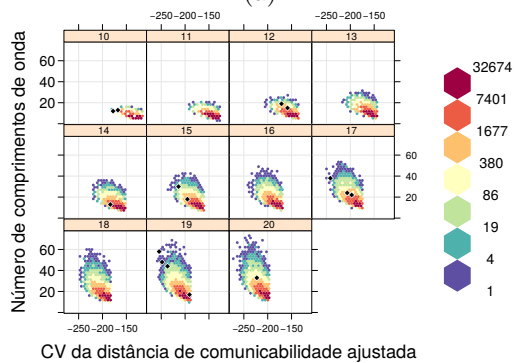
Alguns invariantes apresentam uma tendência relativamente clara da variação da informação mútua em função da ordem n da rede, como os invariantes $k = 1, 12, \text{ e } 14$, mostrados respectivamente nas Figuras 9c, 12c, e 13c. Contudo, alguns invariantes não demonstram um comportamento previsível nesse sentido, como por exemplo, a distância média ($k = 7$), cuja informação mútua versus n pode ser vista na Figura 14a.



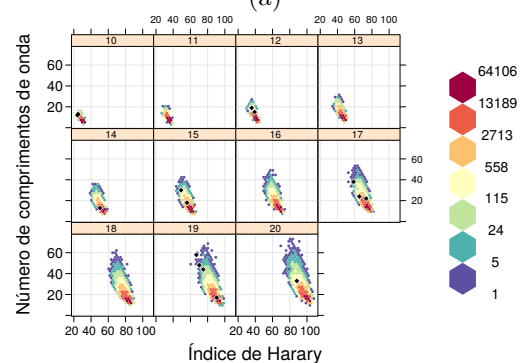
(a)



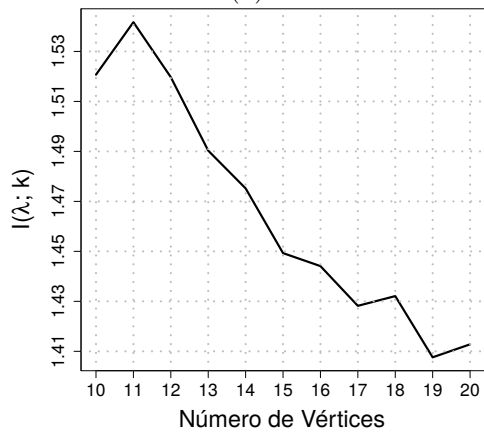
(a)



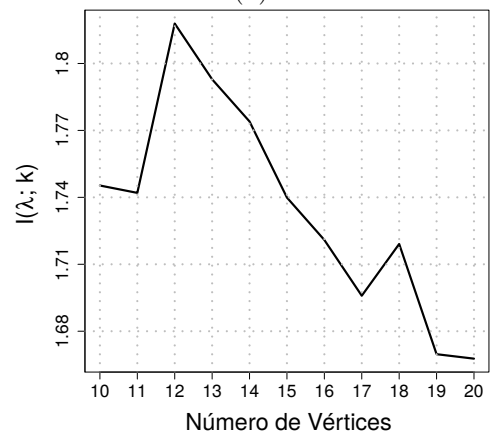
(b)



(b)



(c)



(c)

Figura 12 – Coeficiente de variação da distância de comunicabilidade ajustada ($k = 12$) versus λ para todas as redes do mundo real (representadas por losangos) e redes aleatórias em estudo: redes com todas as ordens n juntas em (a), e separadas pela ordem n em (b). Em (c), é dada a informação mútua do coeficiente de variação da distância de comunicabilidade ajustada e λ , em função da ordem da rede.

Figura 13 – Índice de Harary ($k = 14$) versus λ para todas as redes do mundo real (representadas por losangos) e redes aleatórias em estudo: redes com todas as ordens n juntas em (a), e separadas pela ordem n em (b). Em (c), é dada a informação mútua do Índice de Harary e λ , em função da ordem da rede.

Para alguns invariantes, a informação mútua exibe um intrigante comportamento de “dente de serrate”, onde sua relevância parece depender da paridade da ordem de rede. Um desses invariantes é a transmissão mediana ($k = 8$), onde pode ser visto na Figura 14b que valores maiores de $I(\lambda, k)$ ocorrem para n par. Como consequência, no projeto de topologias de redes ópticas, é mais importante considerar a transmissão mediana para redes com n par do que para redes com n ímpar. A razão para esse comportamento não é clara e requer análise adicional.

É possível verificar - também no Apêndice D - como $I(\lambda, k)$ varia, em função de n , para cada invariante k das Tabelas 2 e 3.

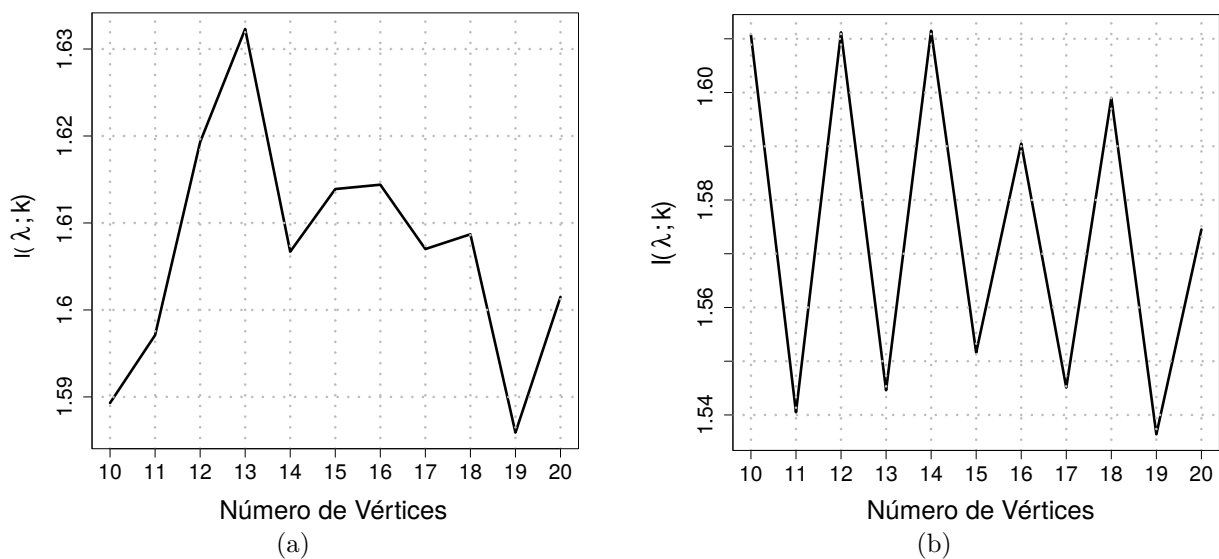


Figura 14 – Informação Mútua da (a) distância média ($k = 7$) e da (b) transmissão mediana ($k = 11$), ambos com λ e em função da ordem da rede (n).

2.3.4 Contrastando os resultados com os achados da literatura

Nas Seções 2.3.1 e 2.3.2 são analisados os invariantes mais importantes com respeito ao requerimento de comprimentos de onda, para todas as redes do mundo real e redes aleatórias em estudo, com as redes de todas as ordens n juntas e para redes com cada n analisado separadamente, respectivamente. Nesses dois casos, é interessante notar que alguns invariantes por vezes aclamados na literatura (vistos na Seção 1.4) não ressaltam no presente estudo com a relevância talvez esperada. Então, na presente seção, são apresentadas análises para esses invariantes, de maneira a observar como eles se encaixam no contexto do presente estudo.

Em Baroni e Bayvel (1997), é afirmado que o número de comprimentos de onda médio não depende do número de nós n da rede, quando se analisa a densidade de arestas (chamada de conectividade física no referido trabalho). Entretanto, como discutido na Seção 2.3.1, o número de vértices ($k = 28$) é, na verdade, relevante. Ainda em Baroni e

Bayvel (1997), verifica-se que a densidade de arestas é indiretamente relacionada a λ médio. No presente trabalho, tal relação não aparenta ser comparativamente tão forte, uma vez que a densidade de arestas ($k = 65$) não aparece na Tabela 2 e nem na Tabela 3. Contudo, a declaração de correlação negativa de Baroni e Bayvel (1997) sobre a densidade da arestas pode ser observada ao analisar visualmente, de modo mais geral, todas as redes juntas na Figura 15a. A Figura 16 mostra os dados da Figura 15a, porém reforçando a informação de que a ordem n da rede importa no requerimento de comprimentos de onda, quando se analisa a densidade de arestas, ao contrário do que diz em Baroni e Bayvel (1997). Por sua vez, analisando mais de perto, para redes separadas por número de vértices na Figura 15b, a força dessa relação se dissipa. Vê-se que essa relação forte observada na Figura 15a é então causada pela combinação de relações pouco definidas e com comportamentos diversos para cada n , apresentadas na Figura 15b.

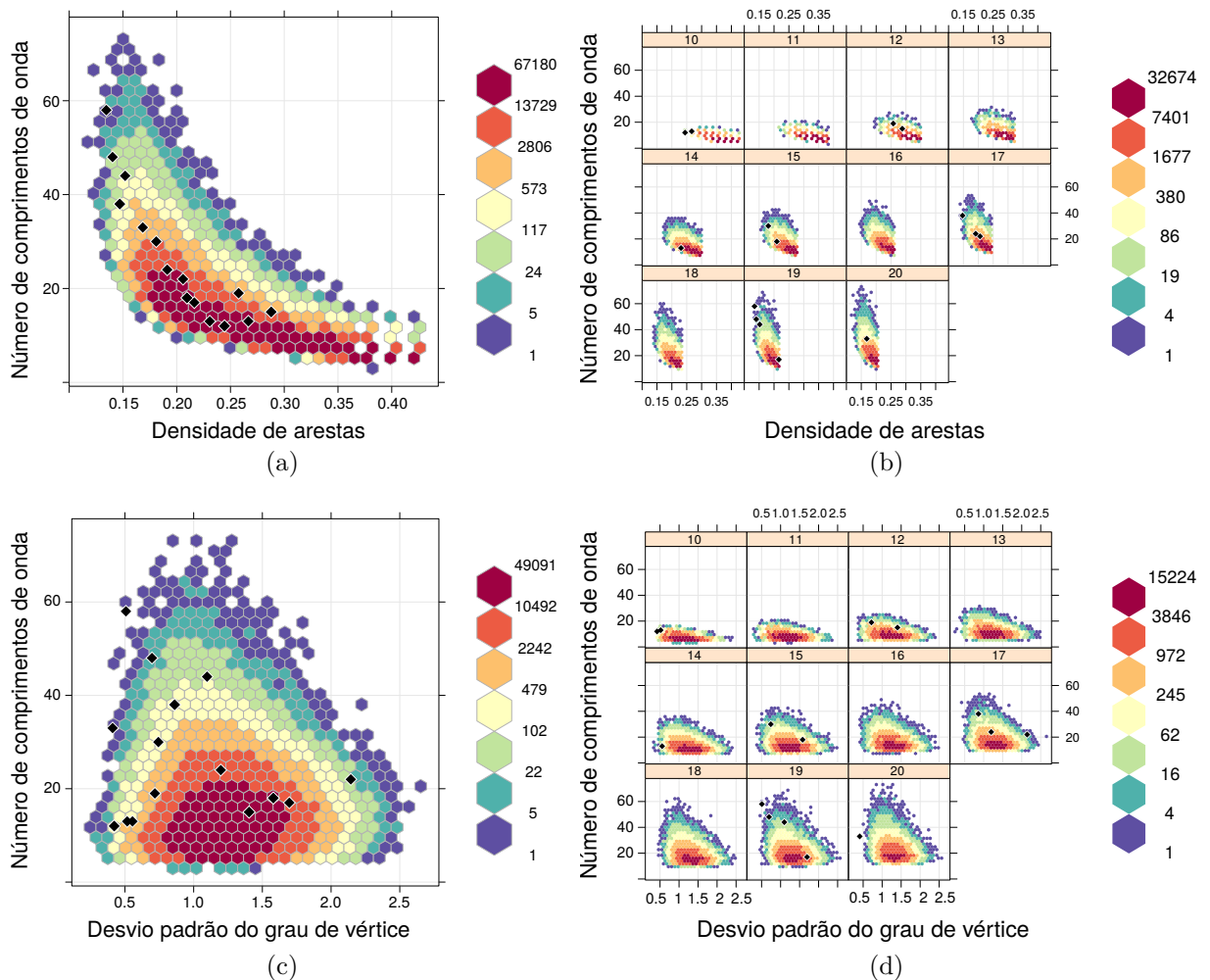


Figura 15 – Densidade de arestas ($k = 65$) e desvio padrão do grau ($k = 211$) ambos versus λ , para as redes do mundo real (representadas por losangos) e redes aleatórias em estudo: para todas as redes de todas as ordens n juntas em (a) e (c), e separadas pela ordem n em (b) e (d).

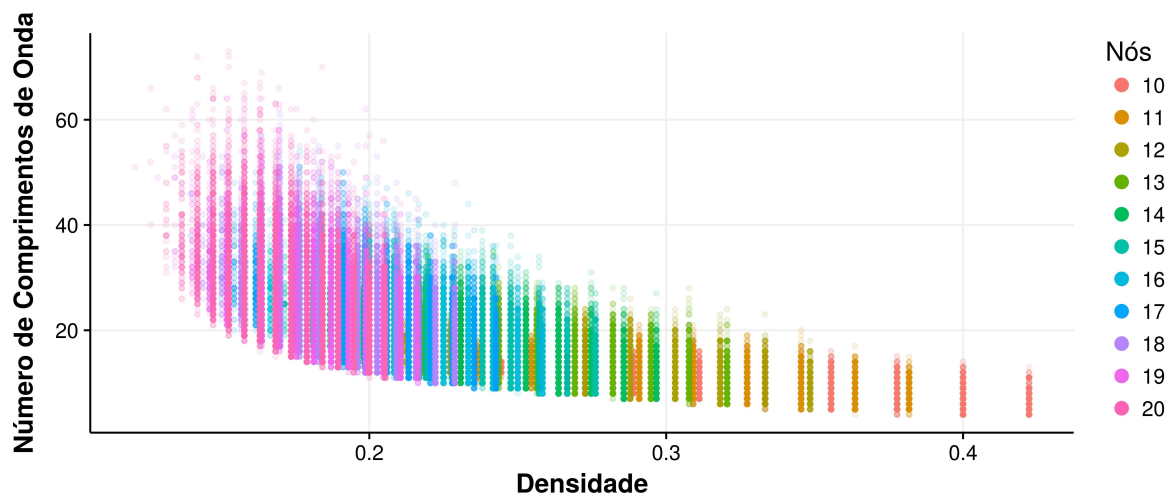


Figura 16 – Densidade de arestas ($k = 65$) versus λ para redes aleatórias em estudo de todas as ordens n juntas.

De acordo com o estudo de [Fenger et al. \(2000\)](#), existe uma relação de correlação positiva entre a variância do grau e λ . Em contraste, tanto em todas as redes reais e aleatórias juntas (Figura 15c) quanto nas separadas pela ordem n (Figura 15d), não se observa tal relação. É necessário entretanto dizer que em vez de analisar a variância do grau, o presente trabalho considera o desvio padrão do grau ($k = 211$), e uma vez que variância e desvio padrão são estritamente ligados, isso não afeta significativamente as análises.

Ainda em [Fenger et al. \(2000\)](#), é constatada uma correlação negativa entre grau médio e λ . Porém, esse comportamento não é claramente detectado quando se observa todas as redes aleatórias e reais juntas (na Figura 17a) ou separadas por n (na Figura 17b). Ao se analisar os valores de $I(\lambda; k)$ das redes separadas por n , o grau médio ($k = 167$) consta em todas as colunas das Tabela 3, mas com baixa importância. Assim sendo, pode-se ser dito que a importância do grau médio não é tão alta quanto se possa supor, pelo menos em comparação aos invariantes k que possuem maior valor de $I(\lambda; k)$. Além disso, em [Fenger et al. \(2000\)](#) é estabelecida uma forte relação entre o número de árvores geradoras e o número de comprimentos de onda. Entretanto, como discutido na Seção 2.3.2, o número de árvores geradoras ($k = 109$), na verdade, não é tão importante quanto o esperado.

De acordo com [Châtelain et al. \(2009\)](#), a conectividade algébrica é indiretamente relacionada com λ . Como discutido na Seção 2.3.2, os resultados sobre conectividade algébrica ($k = 40$) suportam essa descoberta, e adicionalmente esta relação pode ser observada nas Figuras 17c e 17d. Observa-se, entretanto, que as relações dadas não têm o mesmo comportamento, onde a estrutura fornecida na Figura 17c aparenta ser uma composição das relações da Figura 17d.

Como observado no estudo de [Yuan e Xu \(2010\)](#), a distância média é diretamente relacionada com λ em redes *small-world* e *scale-free*. As Figuras 17e e 17f indicam

que essa relação também existe para redes reais e aleatórias do presente estudo, sendo mais forte para as redes reais, eventualmente causada pela menor chance de perceber comportamentos extremos, isso em função do reduzido tamanho dessa amostra, como já foi dito anteriormente. Além disso, o invariante distância média ($k = 7$) é relevante para as redes com todas as ordens n juntas, como visto na Tabela 2, e também para redes com cada ordem n separadamente, como visto na Tabela 3.

2.3.5 Observações adicionais em relação ao projeto de topologias de redes ópticas

Primeiramente, enfatiza-se o fato de que a modelagem de grafos aleatórios considerada representa bem as topologias de redes ópticas do mundo real. De fato, como pode ser visto nas Figuras 9a, 9b, 12a, 12b, 13a, 13b, 15, e 17, ambas as redes reais e as aleatórias em estudo têm um comportamento semelhante em relação a λ . Esse resultado confirma, de alguma forma, a suficiência das hipóteses consideradas, na Seção 2.1.1, para gerar as redes aleatórias.

Em todos os resultados apresentados, notadamente tem-se um mesmo intervalo de requerimento de comprimentos de onda para cada invariante. Isso significa que, para um valor fixado de um dado invariante, é possível encontrar uma topologia de rede que atinge o λ mínimo nesse intervalo. Por exemplo, para cada valor fixado do grau médio (i.e., fixados n e m) na Figura 17b, todas as redes reais e aleatórias com o mesmo grau médio se encontram na mesma linha vertical. Para uma rede real fixada, todas as redes abaixo desta linha vertical requerem menos comprimentos de onda, mantendo os mesmos n e m . Portanto, dados n e m , há espaço para o projeto de novas topologias de redes ópticas que requerem menos comprimentos de onda.

Consolida-se aqui que os invariantes mais relevantes são aqueles que aparecem em todas as colunas da Tabela 3, os quais são listados na Tabela 5. Esses invariantes são classificados nos seguintes grupos temáticos: “Congestionamento” (invariantes que são relacionados com o quanto nós e enlaces são usados para atender a uma demanda de tráfego), “Conectividade” (invariantes que estão relacionados com o número de caminhos disjuntos, por vértices ou arestas, a cada par de vértices), “Distância” (invariantes derivados do conceito de distância entre vértices), e “Grau” (invariantes derivados do grau de vértice), totalizando 20 invariantes topológicos.

Na Tabela 5 há ainda um último grupo chamado de “Aumento da ordem de rede”, que reúne os invariantes relevantes no crescimento da ordem n de rede. Em cada grupo, os invariantes são listados por ordem de relevância.

Dois invariantes altamente correlacionados não são necessariamente redundantes, desde que eles possam de fato fornecer diferentes informações (BENNASAR; HICKS;

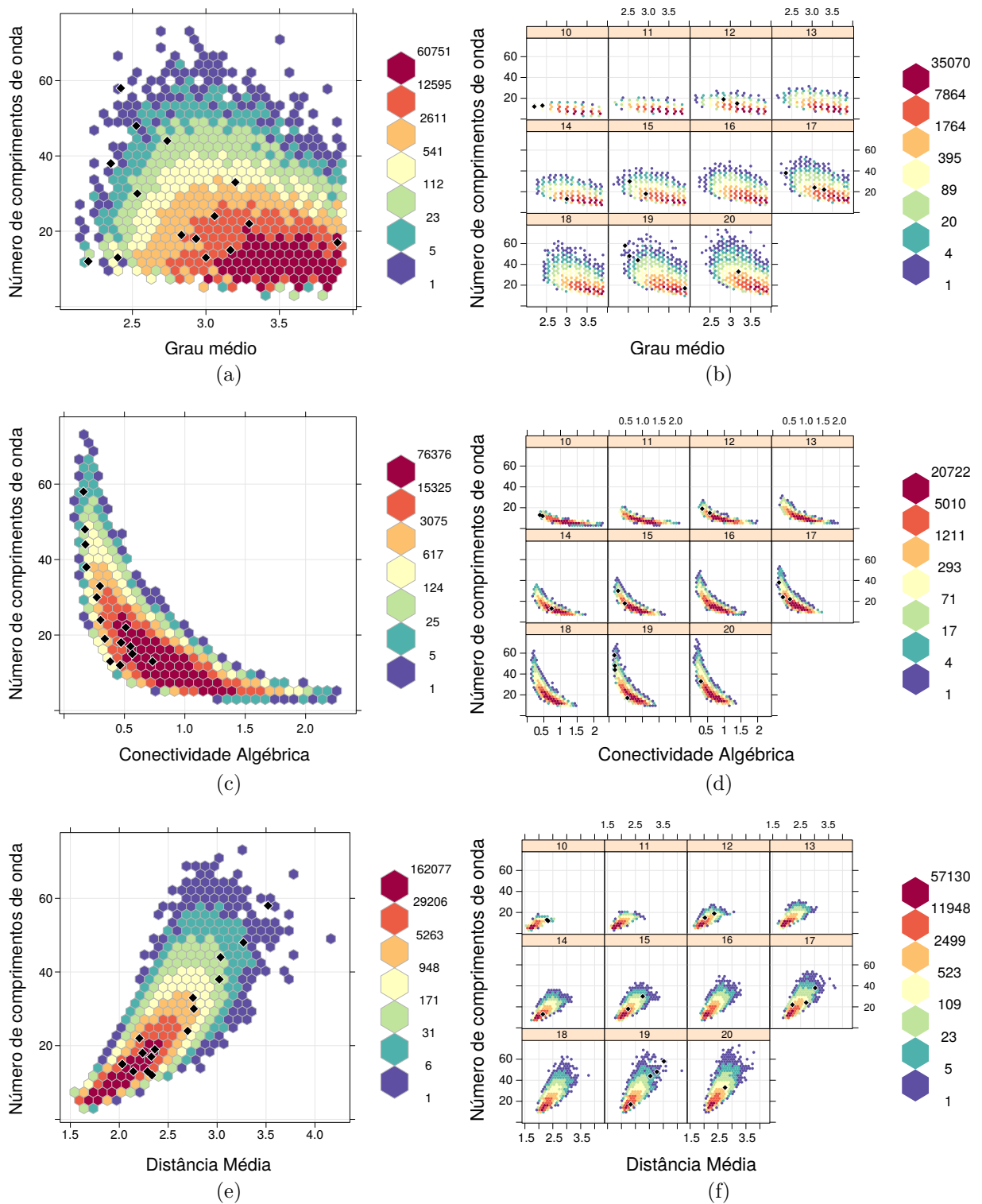


Figura 17 – Grau médio ($k = 167$), conectividade algébrica ($k = 40$) e distância média ($k = 7$), todas versus λ para todas as redes do mundo real (representadas por losangos pretos) e redes aleatórias em estudo: para redes de todas as ordens n em (a), (c), e (e), e para redes separadas por n em (b), (d), e (f).

SETCHI, 2015). Não é recomendado, porém, usar todos os invariantes da Tabela 5, porque isso tornaria a modelagem mais cara e redundante, a menos que se proceda com um método de *feature extraction* (GUYON; ELISSEEFF, 2006). Além disso, dentro de cada grupo, a correlação entre pares de variáveis tende a ser forte em demasia, uma vez que refletem o mesmo tema. Em um projeto de topologia de rede óptica, para cobrir todos os aspectos mencionados na Tabela 5, é recomendado considerar ao menos um invariante de cada grupo temático.

Destaca-se, por fim, o invariante mais proeminente neste estudo, que é o *edge betweenness* máximo, tanto para todas as redes juntas, quanto para as redes separadas por n . Isso poderia ser explicado pela relação existente entre o *edge betweenness* máximo e o congestionamento de aresta. Como observado no estudo de Cousineau *et al.* (2015), o congestionamento de aresta é um *lower bound* para λ , o qual é frequentemente alcançado. Portanto, com estas considerações, é esperada de fato uma forte relação entre o *edge betweenness* máximo e o λ .

Tema	k	Invariante
Congestionamento	1	<i>Edge betweenness</i> máximo
	2	Desvio Padrão do <i>edge betweenness</i>
	3	Amplitude do <i>edge betweenness</i>
	5	<i>Edge betweenness</i> médio
	6	<i>Vertex betweenness</i> médio
Conectividade	40	Conectividade Algébrica
	59	Número de conjuntos mínimos que desconectam o grafo
Transmissão e Distância	4	Índice de Kirchhoff
	7	Distância média
	8	Transmissão média
	9	Transmissão máxima
	10	Índice de Wiener
	11	Transmissão mediana
	12	CV da distância de comunicabilidade ajustada
	14	Índice de Harary
	16	Excentricidade média
	45	Desvio padrão da distância
Grau	101	Mínimo do grau médio dos vizinhos mais próximos
	167	Grau médio
Aumento da ordem de rede	35	Coefficiente de variação de <i>inverse log-weighted similarity</i>
	57	Coefficiente de variação mediano de <i>dice similarity</i>
	61	Coefficiente de variação mediano de <i>jaccard similarity</i>
	63	Coefficiente de variação mediano de <i>inverse log-weighted similarity</i>
	68	Coefficiente de variação mediano de <i>cocitation coupling</i>

Tabela 5 – Invariantes mais relevantes para explicar o número de comprimentos de onda (k é o identificador do invariante). O apêndice A exibe uma breve descrição e o modo de cálculo desses invariantes.

2.4 Conclusão

Após uma série de análises com base na estimação da informação mútua de λ com cada invariante considerado (tanto com todas as redes juntas, quanto com as redes separadas pela ordem n), e de comparativos com a literatura, chega-se a uma seleção de invariantes (dados na Tabela 5), que consistem basicamente de invariantes de congestionamento, conectividade, distância e grau. Tais invariantes são os mais indicados a serem empregados na busca de topologias físicas de redes que minimizem o requerimento de comprimentos de onda, como é o caso da aplicação exibida no Capítulo 3 a seguir.

3 Modelagem via funções substitutas para a busca de redes com menores valores de λ

No presente capítulo são executadas aplicações dos resultados do Capítulo 2. Primeiramente, utilizando-se dos invariantes que melhor explicam o requerimento de comprimentos de onda, funções substitutas são obtidas por meio de regressões apropriadas, para estimativas mais rápidas de λ . Em seguida heurísticas são implementadas, onde as funções substitutas são utilizadas para uma busca mais otimizada de topologias físicas de redes ópticas, e resultados das instâncias executadas mostram uma economia expressiva no requerimento de comprimentos de onda das redes de saída.

Para projetar topologias de redes ópticas com menor requerimento de comprimentos de onda (λ) possível, escolheu-se usar funções substitutas (do inglês, “*surrogate functions*”) computacionalmente mais rápidas para estimar o λ por meio de invariantes topológicos, ao invés de seu cálculo exato (que é, como já foi falado, um problema *NP-hard*), guiando assim o processo de obtenção da topologia da rede de modo um pouco menos preciso, porém mais ágil. No presente capítulo, tais funções substitutas são obtidas por meio de regressões apropriadas, bem como são propostos métodos de otimização que usam essas funções para encontrar (em um intervalo de tempo muito curto) topologias de redes que minimizam o requerimento de comprimentos de onda.

De modo geral, associando o que foi executado no Capítulo 2 com o que é implementado no presente capítulo, é possível sumarizar como segue. Dadas as definições do tipo de rede que se deseja implementar, uma amostra de topologias de rede é criada com restrições que atendam tais definições, de modo a propiciar um cenário alvo de projeto de redes o qual impõe condições topológicas a todas as topologias na amostra. Tais topologias devem ser geradas de modo a estarem dentro de um intervalo grande o suficiente para conter redes com todos os números possíveis de nós que a topologia desejada pode ter. Mais ainda, tal amostra deve conter topologias em quantidade suficientemente grande para garantir a significância estatística das regressões que serão aplicadas nos dados extraídos das topologias da amostra, regressões estas que fornecem as funções substitutas a serem utilizadas na estimação de λ . Para esse fim, tal procedimento baseia-se aqui no cenário alvo de projeto de redes já devidamente tratado na Seção 2.1.1.

Os invariantes de tempo polinomial que já demonstraram (no Capítulo 2) serem mais significativos para explicar λ , bem como o próprio número de comprimentos de onda (λ) que foi calculado de forma exata, terão seus valores usados como dados de entrada nas regressões. No projeto de topologias físicas de redes ópticas aqui proposto, não são consideradas matrizes de tráfego ou restrições de capacidade dos enlaces, dada a abordagem

aqui adotada (Seção 2.1.1), em que o foco é totalmente voltado para obter topologias que minimizem λ baseando-se apenas em parâmetros topológicos da rede, conforme consta no Objetivo Geral (Seção 1.5.1).

A Seção 3.1 trata das regressões estatísticas realizadas para fornecer as equações de estimação de λ , usando os valores obtidos dos invariantes explicativos selecionados e dos valores da variável resposta λ . Finalmente, modelos de otimização heurísticos são dados na Seção 3.2, os quais aplicam as equações de estimação de λ , obtidas via regressão, como funções substitutas na busca de topologias de redes que minimizem λ .

Proceder com uma otimização baseada em uma função substituta emerge do conceito da já fundamentada Metodologia da Superfície de Resposta, do inglês “*Response Surface Methodology*” (RSM), que explora as relações entre várias variáveis explicativas e uma ou mais variáveis resposta (KHURI; MUKHOPADHYAY, 2010). Dessa forma, no presente trabalho busca-se estimar a variável resposta de interesse (λ , no caso) como função de variáveis explicativas dadas por invariantes topológicos de grafo de tempo polinomial, ou como pode-se simplesmente chamar “invariantes explicativos”, os quais são selecionados dentre os mais bem correlacionados com λ , do grupo de invariantes analisado no Capítulo 2. Tais estimadores podem avaliar λ de modo mais eficiente com relação ao tempo do que numa abordagem de um problema linear inteiro, ou de “*integer linear problem*” - ILP, em inglês. Essa avaliação mais rápida permite testar uma quantidade maior de topologias, permitindo explorar o projeto de topologias de redes ópticas em uma escala maior do que com o cálculo exato de λ .

Um resumo da metodologia completa usada, na presente tese, para obter redes otimizadas quanto ao requerimento de comprimentos de onda consiste basicamente de seis passos, os quais podem ser visualizados na Figura 18. Os passos de 1 a 5 cobrem o processo de estimação de λ , onde os passos de 1 a 4 foram já tratados e executados no Capítulo 2, e o passo 5 é tratado na Seção 3.1, o qual usa os resultados fornecidos pelos passos de 1 a 4 para fornecer, por meio de regressões apropriadas, equações para estimação mais rápida de λ em função de variáveis explicativas selecionadas.

No passo 6, as equações de estimação - obtidas no passo 5 - são usadas como funções objetivo substitutas em heurísticas, apresentadas na Seção 3.2, em que se busca obter topologias físicas com as mesmas restrições topológicas do conjunto S do passo 1, mas com valores otimizados da variável resposta λ . O passo 6 consiste então de uma “*Surrogate-Based Optimization*” (QUEIPO *et al.*, 2005), ou seja, uma otimização baseada em funções substitutas ou de aproximação, funções estas que no caso são as equações obtidas pelas regressões no passo 5. Todo esse processo pode ser estendido e aplicado para qualquer outra métrica de redes também difícil de calcular. Na Seção 3.3, são fornecidos resultados obtidos pela aplicação dessas heurísticas, onde são apresentadas as redes alternativas às redes reais analisadas, obtidas pela melhor heurística dentre as propostas.

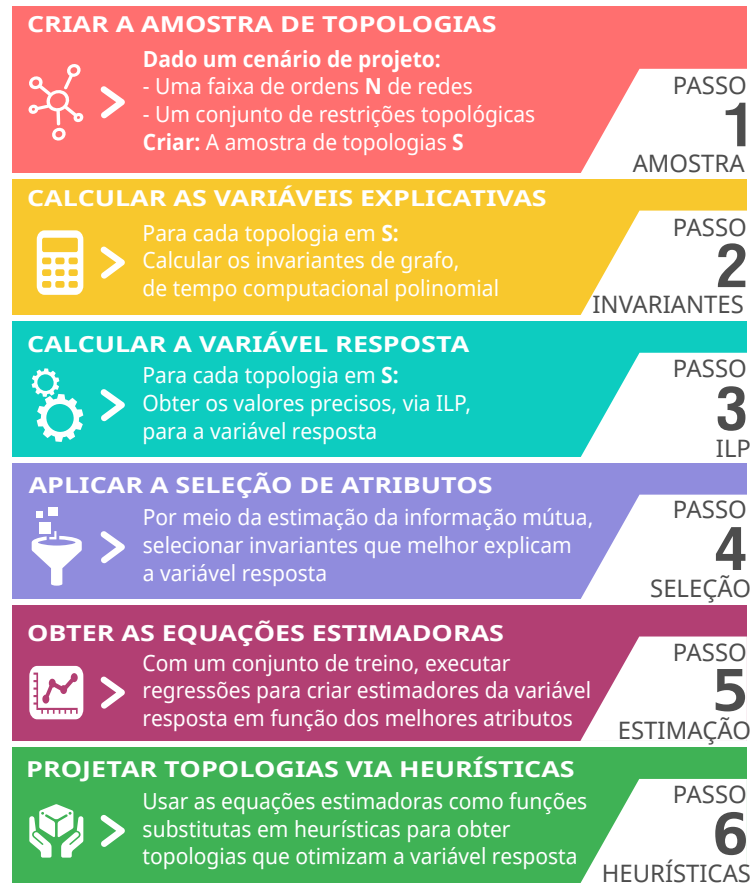


Figura 18 – Etapas do processo para obtenção de topologias com baixo requerimento de comprimentos de onda (λ), baseado numa otimização que usa funções substitutas.

3.1 Obtenção das funções substitutas estimadoras de λ

Na presente seção, são executadas as regressões apropriadas para obtenção das funções substitutas estimadoras de λ , conforme prevê o passo 5 da Figura 18. Para a obtenção dessas funções, sugere-se que a amostra S no passo 1 considere uma faixa de grau médio nas restrições topológicas, assim como foi feito no Capítulo 2, isso para garantir que as topologias estejam conectadas o suficiente, mas também que cada rede não possua enlaces em demasia, de modo que se tornem “caras”, buscando assim uma estimativa de λ o mais realista possível. Desse modo, o número de enlaces m das topologias varia mesmo com $n \in N$ fixo, permitindo avaliar mais possibilidades, caso seja desejado ou necessário.

O tamanho da amostra de grafos S , usada para obtenção dos dados para executar as regressões apropriadas, é bem menor do que o tamanho considerado para o ranqueamento dos invariantes no Capítulo 2, isso para buscar otimizar a execução dos testes das regressões, mas ainda garantindo resultados com significância estatística. Então, sobre o tamanho da amostra de treino S considerada para as regressões, nós chamamos de NS o número de grafos tomados para cada $n \in N$ (lembrando que n é a ordem do grafo, e N é o vetor

com todas as ordens consideradas). Com um tamanho de amostra $|S| \geq 2847$, e logo com $NS \geq 2847/|N|$, o poder do teste em Modelos Lineares Generalizados (COHEN, 1987) garante uma confiança de 99%, com um poder estatístico de 98% (vide Apêndice E). Para executar as regressões utiliza-se então um subconjunto dos grafos escolhidos aleatoriamente dentre os considerados no Capítulo 2, bem como são aproveitados os cálculos dos invariantes já realizados para cada grafo. Como foi visto, tais grafos variam entre 10 e 20 nós, e com isso temos que $|N| = 11$. Sendo assim, por simplicidade tomamos $NS = 300$, e com isso temos $|S| = 3300$, o que atende a restrição $NS \geq 2847/|N|$.

Para decidir o tipo de regressão a ser utilizada, na amostra S (de tamanho $|S| = 3300$) foi analisado o histograma da variável resposta de interesse λ , que é uma variável de contagem. Foi verificado que ela aparenta ter uma distribuição Poisson, como pode ser visto na Figura 19. Então, com isso, a regressão de Poisson (HAIGHT, 1967) aparenta ser adequada para estimar o λ esperado, $E[\lambda]$, no passo 5. A regressão de Poisson faz parte da já citada classe de Modelos Lineares Generalizados (do inglês, *Generalized Linear Models* - GLM) (FOX, 2015).

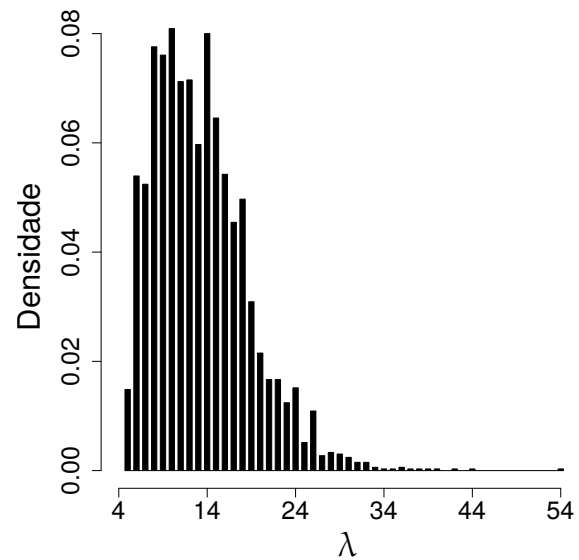


Figura 19 – Histograma de λ das redes de S .

Ainda no passo 5, uma regressão quantílica (KOENKER, 2005) é escolhida para estimar o valor mínimo que λ poderia assumir, denotado por $E_{min}[\lambda]$. Enquanto muitas regressões que usam o método dos mínimos quadrados, como a própria regressão Poisson, produzem estimativas próximas da média, uma regressão quantílica produz estimativas próximas da mediana ou de qualquer outro quantil de interesse. A escolha da regressão quantílica é aqui tão somente um artifício para se obter estimativas o mais próximas possível dos menores valores plausíveis de λ , escolhendo o menor quantil possível que ainda permita uma significância em tal regressão. Ensaios executados mostraram que um valor de quantil igual a 0,01 é o com menor ordem de grandeza em que a regressão quantílica se mantém significativa, com os dados das topologias utilizados, em testes via p-valor (WASSERSTEIN; LAZAR, 2016). Um quantil igual a 0,01 significa que $E_{min}[\lambda]$ estima valores limitados aos 1% menores valores de λ possíveis para cada configuração de entrada da regressão. Similarmente, a regressão de Poisson também demonstrou significância estatística via p-valor. Assim sendo, as equações a serem estimadas, advindas desses dois

modelos, são dadas por:

$$E[\lambda] = \exp(a_1 + a_2n + a_3m + a_4EB_{max} + a_5AC), \quad (3.1)$$

$$E_{min}[\lambda] = b_1 + b_2n + b_3m + b_4EB_{max} + b_5AC, \quad (3.2)$$

onde os coeficientes estimados a_i e b_i ($i = 1, \dots, 5$) são as saídas das regressões, EB_{max} é o máximo *edge betweenness* e AC é a conectividade algébrica.

Sobre os invariantes explicativos escolhidos para comporem as Equações 3.1 e 3.2, foi imposta uma premissa inicial de que essas equações de estimação de λ incluíssem o número de nós n e o número de enlaces m , que são igualmente considerados invariantes de grafos e estão diretamente relacionados com a topologia da rede projetada (FENGER *et al.*, 2000), de modo que qualquer outra inserção de invariante foi considerada apenas depois dessa inserção inicial de n e m . Essa decisão colabora diretamente com o projeto de topologias de redes ópticas, em que se torna possível ter como entrada esses dois dados a princípio simples, mas de extrema importância na rede.

Em seguida, escolheu-se invariantes dentre os que se destacaram no ranqueamento executado no Capítulo 2 (vide Tabela 5), a saber: o melhor ranqueado na categoria de congestionamento, que foi o *edge-betweenness* máximo (EB_{max}), e o melhor na categoria de conectividade, que foi a conectividade algébrica (AC). O *edge-betweenness* reflete a proporção de menores caminhos que passam por um enlace (LU; ZHANG, 2013), e a conectividade algébrica é o segundo menor autovalor da matriz laplaciana do grafo da topologia (CHUNG, 1997). Foi considerado inserir outros invariantes bem correlacionados com λ , de outras categorias (também dadas na Tabela 5), mas testes com tais inserções mostraram que as regressões (passo 5 da Figura 18) não prosseguiram estatisticamente significantes, isso talvez motivado pela alta correlação entre os invariantes explicativos inseridos e os já presentes na regressão.

Postas todas essas condições, de posse dos valores de λ , n , m , EB_{max} e AC dos grafos da amostra de treino S , executou-se as regressões de Poisson e quantílica no programa R, onde obteve-se uma estimativa dos coeficientes a_i e b_i ($i = 1, \dots, 5$) das Equações 3.1 e 3.2, e as seguintes equações foram obtidas:

$$E[\lambda] = \exp(2,1060 + 0,0445 n - 0,0036 m + 0,0229 EB_{max} - 0,6447 AC) \quad (3.3)$$

$$E_{min}[\lambda] = 1,9233 + 0,8532 n - 0,2310 m + 0,2912 EB_{max} - 3,0782 AC. \quad (3.4)$$

Uma vez executadas as regressões com os valores que os invariantes considerados assumem para cada grafo da amostra S , e então de posse das Equações 3.3 e 3.4, tem-se que dada uma amostra S' (tão grande quanto se queira) de topologias de redes (que atendam às mesmas restrições topológicas de S), pode-se obter os invariantes n , m , EB_{max} e AC dessas redes em S' , e com esses valores é possível então estimar, em tempo polinomial, um

“ λ esperado” (médio) e um “ λ mínimo” possível para as topologias dessa amostra S ’, ao invés de resolver um problema *NP-hard* para obter o valor exato de λ para cada rede.

Chama-se a atenção para o fato de que as Equações 3.3 e 3.4 obtidas são válidas para serem aplicadas apenas para topologias que atendam às restrições topológicas da amostra S . Para redes com quaisquer mudanças nessas restrições, um novo conjunto de treinamento S deve ser obtido, com novas restrições que atendam essas outras redes, e uma nova regressão deve ser executada.

Pode-se notar que todos os sinais dos coeficientes nas Equações 3.3 e 3.4 são consistentes com a correlação esperada de cada invariante topológico das equações com λ . O coeficiente de m é negativo uma vez que quanto maior o número de enlaces na rede, menor é o λ necessário para transportar as demandas de tráfego. Por outro lado, n tem um coeficiente positivo uma vez que o crescimento de n implica em mais tráfego para a rede lidar. EB_{max} tem um coeficiente positivo porque valores maiores desse invariante indicam a existência de enlaces mais congestionados, o que acarreta um maior requerimento de λ . Por fim, como AC é limitado superiormente pela conectividade de vértices (GROSS; YELLEN; ZHANG, 2013), esta convergindo para zero, a rede tende a ficar menos conexa, e implicando ainda que os valores de AC tendam a zero (com $AC = 0$ em topologias desconexas), logo a diminuição da conectividade da rede, que implica em um aumento na demanda de λ , se traduz em diminuição dos valores de AC , o que explica o sinal negativo do coeficiente de AC . A ocorrência de sinais positivos ou negativos nos coeficientes das Equações 3.3 e 3.4 pode ainda ser verificada visualmente nas tendências de cada invariante com λ , apresentadas nos gráficos 9a, 10, 17c e 46b.

Para concluir esta seção, discute-se a utilidade do método proposto em encontrar topologias eficientes em tempos curtos, durante a fase de projeto da rede. Por exemplo, ilustra-se a procura de redes alternativas para a NSFNET. Como essa rede possui $n = 14$ nós e $m = 21$ enlaces, existem então 161.191.179 diferentes topologias 2-conexas com o mesmo n e m , calculado pelo programa *Nauty*. Caso se tentasse exaustivamente procurar a solução ideal que forneça o menor valor exato de λ verificando cada uma dessas milhares de topologias, isso tomaria mais de 8507 horas de tempo de CPU, que é quase um ano (considerando o tempo de 0,19s para o caso NSFNET, dado na Figura 20).

A maior parte do tempo gasto em todos os passos dados na Figura 18 é no cálculo exato de λ , dado no passo 3. A Figura 20 mostra então o tempo gasto apenas nesse passo para algumas redes reais dentro da faixa de N usada nas regressões (linha preta) e outras redes maiores (linha vermelha), que demonstram um crescimento exponencial do tempo computacional em função do tamanho da rede¹. Por outro lado, o tempo para calcular os estimadores de λ (via as equações de regressão) segue insignificante, independentemente do tamanho da rede (linha verde), e essa observação é especialmente importante quando o tamanho da rede cresce além de $n = 20$. Portanto, o uso de estimadores reduz drasticamente

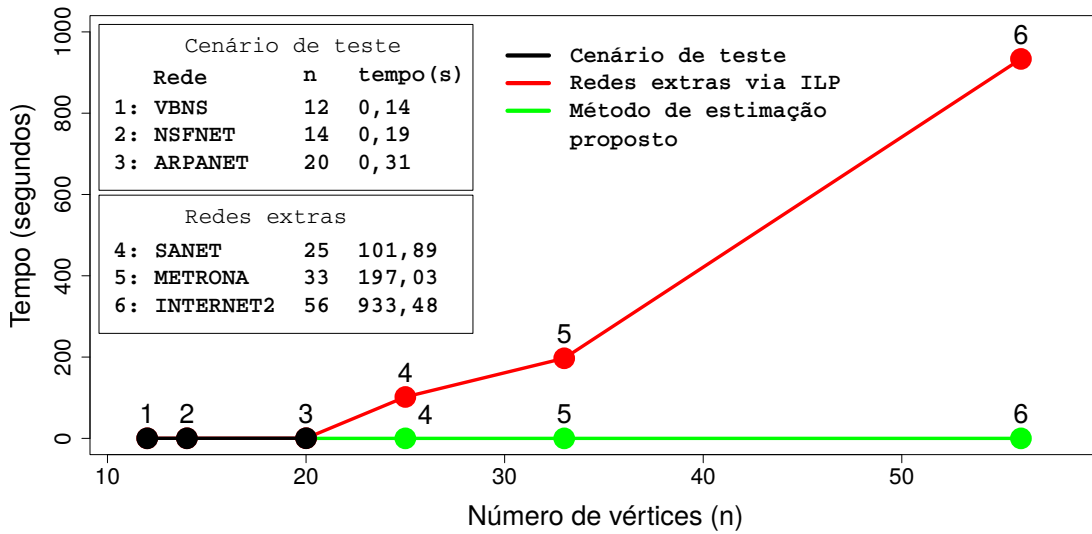


Figura 20 – Tempo de CPU versus o número de nós n para o cenário de teste, o qual compara o método de estimação de λ via as Equações 3.3 e 3.4 (em verde) e a solução exata (em vermelho). Dados de tempo computacional gasto na solução exata obtidos em Cousineau *et al.* (2015).

o tempo computacional, ainda que não forneça necessariamente a solução exata. Esse efeito ocorre porque a estimação de λ usa equações de tempo polinomial, em que a solução exata é aplicada apenas na amostra de treinamento S (para gerar as equações de regressão), onde com $NS = 300$ e $N = 10, 11, \dots, 20$, a amostra S com 3300 topologias precisa de cerca de 12 minutos de tempo de processamento para calcular o λ exato para cada elemento de S (considerando um tempo médio de 0,19s para cada rede).

Na Seção 3.1.1 dada a seguir, testes práticos são executados com as equações das regressões obtidas na presente seção, dentre eles o quão perto os valores reais de λ estão de suas estimativas.

3.1.1 Teste de validação

A presente seção busca demonstrar o desempenho da estimação de λ por meio das Equações 3.3 e 3.4 obtidas pelas regressões. Para medir a efetividade dessas equações, elas são testadas em uma segunda amostra também contendo 3300 topologias diferentes e aleatórias, diferentes ainda das da primeira amostra usada na regressão, mas com as mesmas restrições topológicas e também com $n = 10, \dots, 20$. Para cada topologia dessa segunda amostra são contrapostas as estimativas $E[\lambda]$ e $E_{min}[\lambda]$ com o valor exato de λ .

A correlação linear entre λ exato e $E[\lambda]$ é 97,3%, que é considerada uma correlação muito forte, o que estabelece que a estimativa $E[\lambda]$ informa bem sobre os valores exatos de λ . Diante disso e de outros bons resultados de estimação alcançados, como os mostrados

¹Tempo computacional na Figura 20 refere-se a uma máquina Intel Xeon Processor E5-2430 v2 com 96 GB DDR3 de RAM.

na Tabela 6 apresentada a seguir, mesmo que testes de hipótese tenham dado negativo a respeito de λ seguir exatamente uma distribuição de Poisson, optou-se por seguir com a regressão Poisson, dada a sua capacidade demonstrada em estimar λ . Assim, é realizada uma regressão linear com $E[\lambda]$ explicando λ ($\lambda = a \cdot E[\lambda] + c$), que obteve um coeficiente angular $a = 0,99$ e um intercepto $c = 0,17$. Um coeficiente angular próximo de 1 e um intercepto próximo de zero confirmam a boa correlação. A Figura 21a ilustra essa correlação, em que os pontos azuis estão relativamente perto da linha vermelha, que é a regressão linear entre $E[\lambda]$ e λ . Mais ainda, executando uma regressão linear com $E_{min}[\lambda]$ explicando λ , são encontrados um intercepto de $-1,33$ e um coeficiente angular de $1,36$ (significativamente maior que 1), o que significa que $E_{min}[\lambda]$ de fato tende a ser menor que λ , como esperado (Figura 21b).

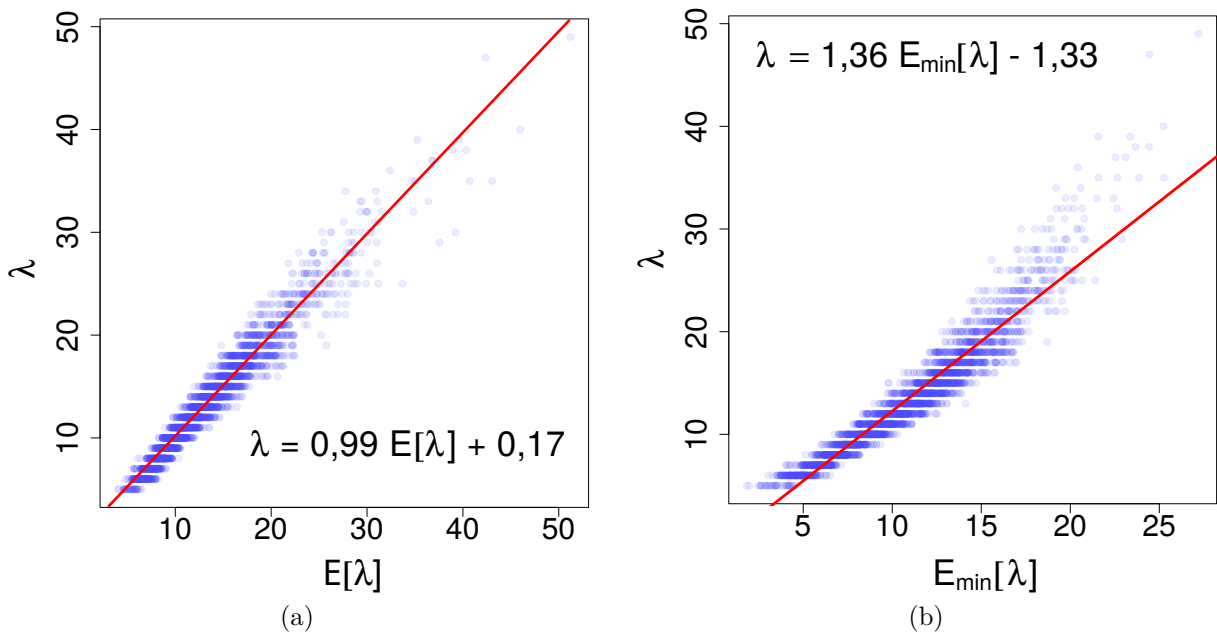


Figura 21 – Visualização da relação linear (em vermelho) entre $E[\lambda]$ e λ em (a), e entre $E_{min}[\lambda]$ e λ em (b). Dados da segunda amostra de 3300 topologias (em azul) em ambos (a) e (b).

Em seguida, especificamente para as topologias com $n = 10, 15$ e 20 da segunda amostra, são calculadas as diferenças

$$D_1 = \frac{\lambda - E[\lambda]}{\langle \lambda \rangle},$$

e

$$D_2 = \frac{\lambda - E_{min}[\lambda]}{\langle \lambda \rangle},$$

e os resultados podem ser vistos na Tabela 6. Note que D_1 e D_2 são porcentagens relativas ao λ médio, simbolizado por $\langle \lambda \rangle$. Na Tabela 6, coluna $|D_1| < 15\%$, vê-se que uma massiva parte da estimacão $E[\lambda]$ difere de λ de no máximo 15% de $\langle \lambda \rangle$, mostrando uma consistência do estimador $E[\lambda]$. Sobre D_2 , na coluna $D_2 < 0$ uma parte muito insignificante

de D_2 assume valores menores que zero. Este comportamento implica que quase todas as estimativas $E_{min}[\lambda]$ são valores menores que λ , o que é o esperado. Mais ainda, a coluna $D_2 < 30\%$ diz que é possível reduzir o requerimento de λ em até 30% de $\langle \lambda \rangle$ em uma grande parcela de casos. E ainda na coluna $D_2 < 30\%$, para $n = 10$ pode ser inferido que até 17% das topologias podem ter uma redução de 30% ou mais no requerimento de λ . Para $n = 15$, essa taxa é de 11,7%, e para $n = 20$ ela é de 21,7%. Finalmente, observa-se que os melhores valores para D_1 e D_2 são obtidos para $n = 15$, o qual é o valor central do intervalo N . Portanto, sugere-se que, se um n inicial é previsto (mas não definido) para uma rede que está sendo projetada, esse n deve ser o valor central no intervalo N , já que nele (e perto dele) estão as melhores estimativas de λ .

Tabela 6 – Resultados das diferenças entre cada λ e seus valores estimados.

	$\langle \lambda \rangle$	$D_1 = (\lambda - E[\lambda]) / \langle \lambda \rangle$			$D_2 = (\lambda - E_{min}[\lambda]) / \langle \lambda \rangle$			
		$ D_1 < 5\%$	$ D_1 < 10\%$	$ D_1 < 15\%$	$D_2 < 0$	$D_2 \geq 0$		
						$D_2 < 10\%$	$D_2 < 20\%$	$D_2 < 30\%$
n = 10	7,1	34,0%	61,3%	80,3%	1,0%	11,0%	50,3%	83,0%
n = 15	13,3	44,3%	78,0%	94,3%	0,0%	25,0%	67,0%	88,3%
n = 20	19,9	44,0%	76,0%	90,7%	1,6%	20,3%	57,7%	78,3%

3.2 Propositura de heurísticas para a busca de redes com menor requerimento de λ via funções objetivo substitutas

Para ilustrar o que pode ser feito com os resultados produzidos até aqui, na presente seção são propostas e implementadas heurísticas genéricas e básicas da literatura, voltadas aqui para o projeto de topologias físicas de redes ópticas, baseadas em processos de otimização via funções objetivo substitutas. Para isso, na busca de topologias de redes ópticas que minimizem o requerimento de λ , essas heurísticas utilizam como funções objetivo substitutas os estimadores de λ , de tempo polinomial, obtidos na Seção 3.1. O uso de tais estimadores acarreta numa pequena perda de precisão na estimativa, mas agrega ganho computacional às heurísticas, se comparado ao cálculo exato *NP-hard* de λ .

São propostas aqui quatro heurísticas para a busca de topologias, cada uma baseada nas seguintes abordagens: (i) busca aleatória com escolha randômica das topologias a serem testadas, (ii) busca local (do inglês *Local Search*) do tipo *Downhill* (algoritmo guloso), (iii) busca local com análise de toda a vizinhança, e (iv) busca em vizinhança variável (do inglês *Variable Neighborhood Search* - VNS) (BURKE; KENDALL *et al.*, 2014). Cada uma dessas heurísticas propostas é detalhada em fluxogramas, dados das Figuras 22 até 25.

Explanando inicialmente sobre a heurística baseada em busca aleatória, dada no fluxograma da Figura 22, vê-se que o número de vértices n e o número de arestas m são entradas obrigatórias do modelo, pois a rede final resultante do processo será uma rede com

n nós e m enlaces, necessariamente dados na entrada. O *goal*, porém, é um parâmetro não obrigatório de entrada, e traz com ele o valor de λ que se objetiva para a rede projetada. Se *goal* não for fornecido, duas situações podem ocorrer: (i) se o parâmetro de entrada *ForceMin* for dado como TRUE, ao *goal* será atribuída a estimativa de $E_{min}[\lambda]$, ou seja, a busca cessa assim que for encontrada uma rede com o menor λ estimado possível, isso se o tempo, dado pelo parâmetro de entrada *stopT*, não finalizar antes, sendo então fornecida a última melhor rede encontrada, e (ii) caso contrário, com *ForceMin* diferente de TRUE, a busca pela topologia não terá limitação inferior para λ , e apenas cessará quando o tempo finalizar. Logo, vê-se que *stopT* é um parâmetro obrigatório, sendo um limitante de tempo a ser usado na execução do algoritmo da heurística.

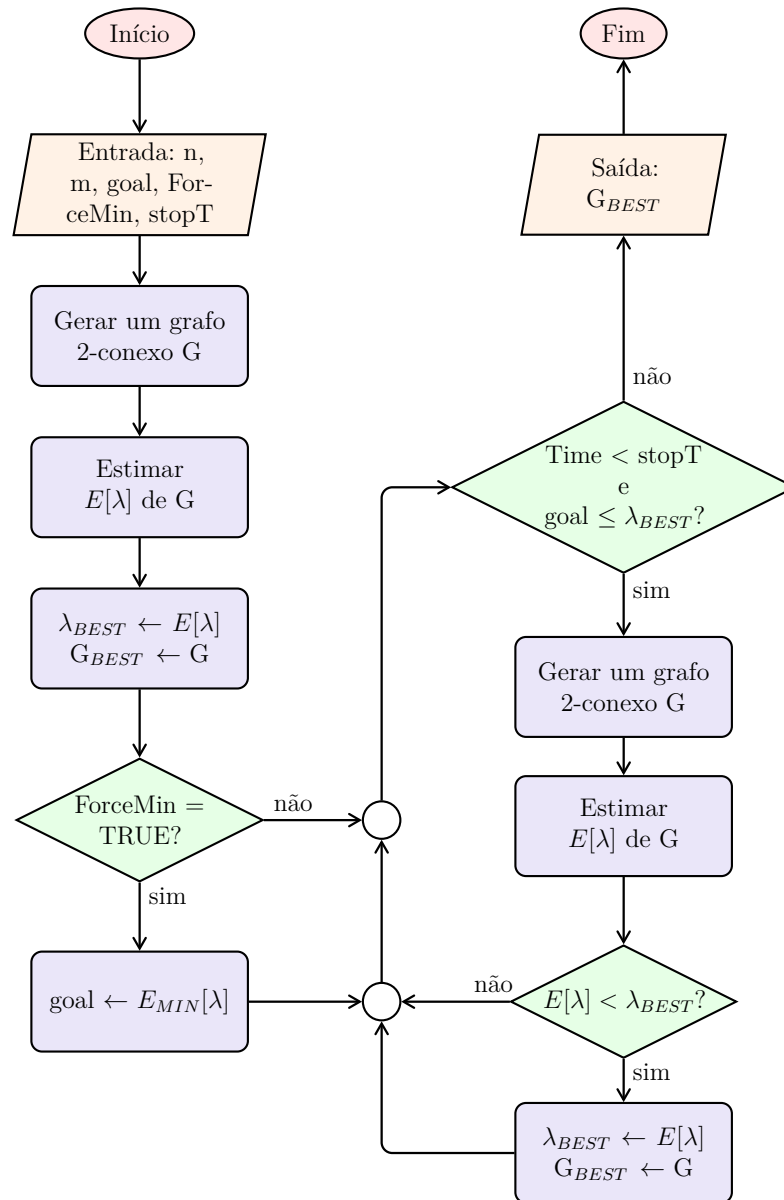


Figura 22 – Heurística baseada na busca aleatória, ou seja, na escolha randômica das topologias a serem testadas.

De modo geral, pode-se dizer que a cada novo grafo tomado aleatoriamente, a heurística de busca aleatória calcula a sua estimativa esperada de λ , $E[\lambda]$, e verifica se esse novo grafo tem estimativa melhor (menor) que o grafo salvo com a menor solução encontrada até então. Se sim, o grafo solução é substituído. Se não, a busca aleatória segue. Assim que λ alcançar o *goal* ou assim que o tempo *stopT* acabar, o processo para, e a última solução encontrada é fornecida como solução final. Cientes de que a busca aleatória de grafos não é eficiente, o objetivo aqui é verificar se as estimativas de λ podem contribuir nesse cenário ineficaz de busca.

Observa-se que na heurística de busca aleatória busca-se gerar apenas grafos 2-conexos para avaliação, isso para que redes com essa característica sejam fornecidas ao fim, lembrando que tal tipo de rede também foi considerada na obtenção das equações de estimação de λ . Todas as heurísticas que são apresentadas na sequência também trabalham nesse sentido.

O passo seguinte à busca aleatória é construir uma heurística baseando-se no conceito de busca local. Ainda sem sofisticar muito, pode-se considerar um método *Downhill* (descida, em inglês), que reflete um algoritmo guloso, onde nesse método busca-se opções de redes em uma vizinhança da rede solução atual e, na varredura dessa vizinhança, assim que é encontrada alguma rede com menor λ esperado, $E[\lambda]$, a solução é atualizada para tal rede e a nova vizinhança passa a ser a vizinhança dessa nova rede, sem inclusive terminar de verificar toda a vizinhança anterior, onde vê-se aí o conceito de *Downhill* e de algoritmo guloso, onde a primeira solução melhorada encontrada já é dada como nova solução naquele dado instante.

A heurística do método *Downhill* é dada no algoritmo da Figura 23. Nela é possível ver que além da entrada do número de vértices n e do número de arestas m , existe alternativamente a opção de se entrar com uma rede inicial, onde nesse caso a busca por soluções se inicia pela vizinhança dessa rede de entrada. Mas se é escolhido entrar com n e m então é gerada uma rede aleatória para iniciar as buscas.

Observa-se que assim como na heurística de busca aleatória, a heurística *Downhill* também tem os parâmetros de entrada *goal*, *ForceMin* e *stopT*, que atuam da mesma forma que na heurística de busca aleatória. Na heurística *Downhill* há ainda um parâmetro a mais de entrada, o *sorted*, que se coloca do seguinte modo: (i) se *sorted* for igual a FALSE, a vizinhança é criada tirando uma aresta do grafo solução atual e inserindo uma aresta aleatória do complementar do grafo, ou seja, a inserção da aresta não é um procedimento ordenado, e toda essa ação é repetida para cada aresta do grafo, gerando assim o conjunto de vizinhança de grafos, e (ii) se *sorted* for diferente de FALSE, a diferença é que a inserção das arestas no complementar do grafo obedece uma ordem definida pela rotulação das arestas. Com esse procedimento, tem-se que cada vizinhança gerada possui m grafos.

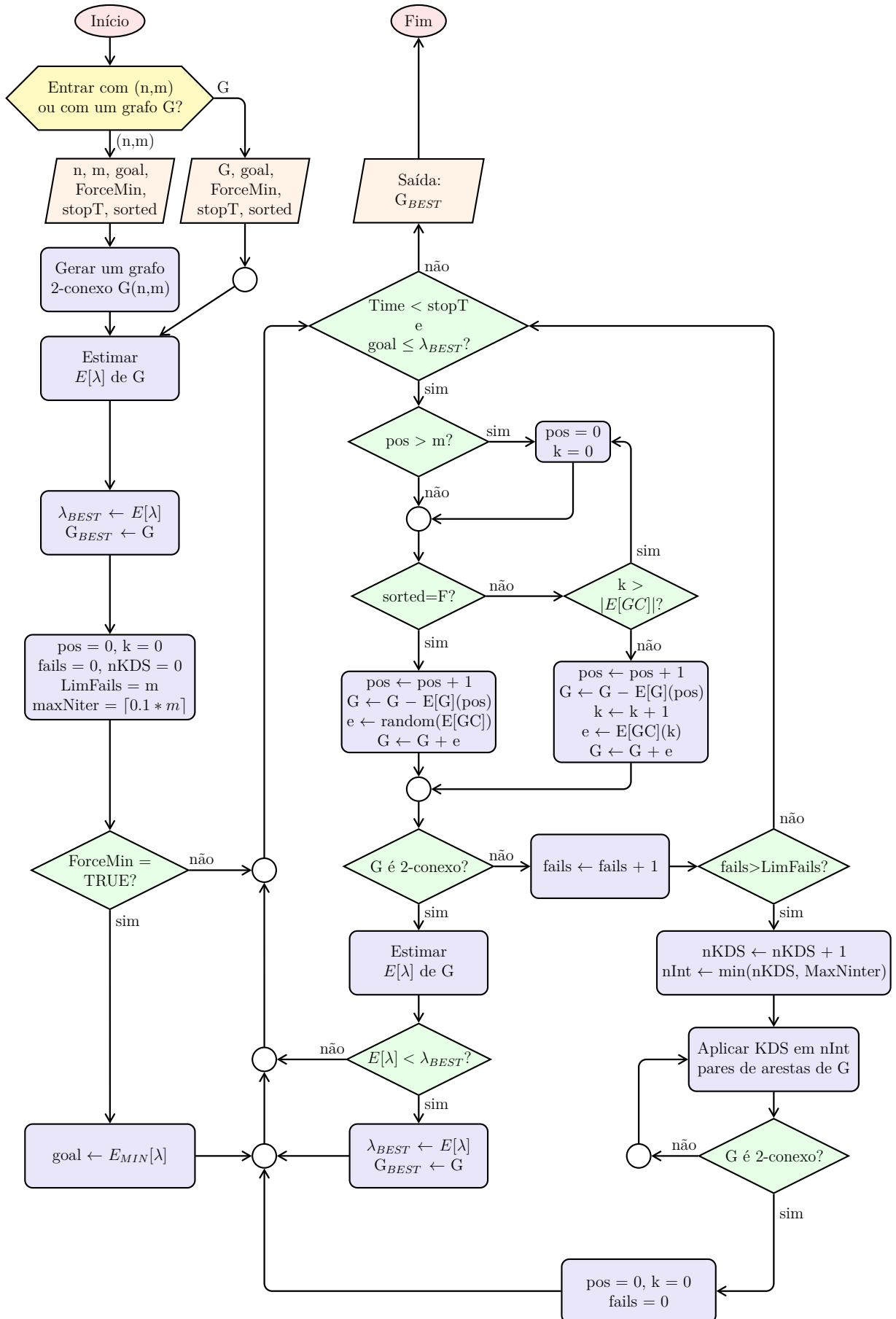


Figura 23 – Heurística de busca local do tipo *Downhill* (algoritmo guloso).

Como a criação da vizinhança de grafos exige a retirada e inserção de arestas, não se tem como ter certeza de como isso afeta a 2-conexidade dos grafos gerados, portanto, buscando controlar essa questão, utiliza-se uma variável de controle intitulada *fails*. Tal variável mensura o quanto grafos não 2-conexos estão sendo gerados, e se a frequência dessa ocorrência for significativa (dentro de um parâmetro escolhido para o modelo) é executado um procedimento que chamamos de KDS, para mudar a vizinhança para uma outra “localidade”, buscando com isso sair de um potencial “buraco” com poucas redes 2-conexas.

O procedimento KDS (*Keeping Degree Sequence*, em inglês) modifica o grafo, preservando a sequência de graus dos vértices, e conseqüentemente também o grau médio, uma vez que n e m não são modificados no processo. Tal procedimento consiste em tomar, do grafo de solução atual, um par aleatório de arestas (a, b) e (c, d) e trocar pelo par (a, d) e (c, b) , se estes últimos já não existirem no grafo (CSARDI; NEPUSZ, 2006).

Aprimorando um pouco a busca local *Downhill*, pode-se considerar investigar toda a vizinhança, ao invés de apenas tomar a primeira melhor solução encontrada. Essa abordagem é dada na heurística da Figura 24, onde é apresentado um algoritmo de busca local com inspeção completa das vizinhanças. Observa-se que nessa nova abordagem, a topologia de rede de início pode se dar tanto com o par (n, m) bem como com uma rede G , assim como é na heurística *Downhill*; e também, da mesma forma trabalham os parâmetros de entrada *goal*, *ForceMin* e *stopT*. Há aqui um parâmetro a mais, denominado *set*, que diz respeito a qual será o tipo de vizinhança a ser testada, onde três tipos de vizinhanças são considerados.

A primeira vizinhança é denominada de S porque é a menor das três (S de *small*, pequeno em inglês). Quando *set* = S na entrada, a vizinhança a ser completamente inspecionada é gerada como no caso *sorted* = FALSE do modelo *Downhill* dado anteriormente, ou seja, é gerada pela retirada de uma aresta do grafo G e insere-se uma aresta escolhida aleatoriamente do grafo complementar, e repetindo esse processo para todas as arestas do grafo G atual, gera-se uma vizinhança de tamanho m (m = número de arestas).

A segunda vizinhança ocorre com *set* = M que é, no caso, a de tamanho médio dentre as três. Para gerá-la, retira-se uma aresta aleatória do grafo G atual e insere-se cada uma das arestas complementares de vez, gerando uma vizinhança com $(n(n - 1)/2) - m$ grafos. Se esse procedimento for repetido para cada aresta do grafo original, tem-se então $m[(n(n - 1)/2) - m]$ grafos, que é a vizinhança no caso de *set* = L (L de *large*, grande em inglês). De todos os grafos gerados em *set* = S, M e L, apenas os 2-conexos são considerados para realmente compor cada uma das três vizinhanças geradas.

Para contornar problemas de não existir vizinhança com pelo menos um grafo 2-conexo, ou ainda, caso ocorram numerosas tentativas sucessivas e fracassadas na busca de uma rede com menor requerimento de λ estimado, é usada uma variável de controle,

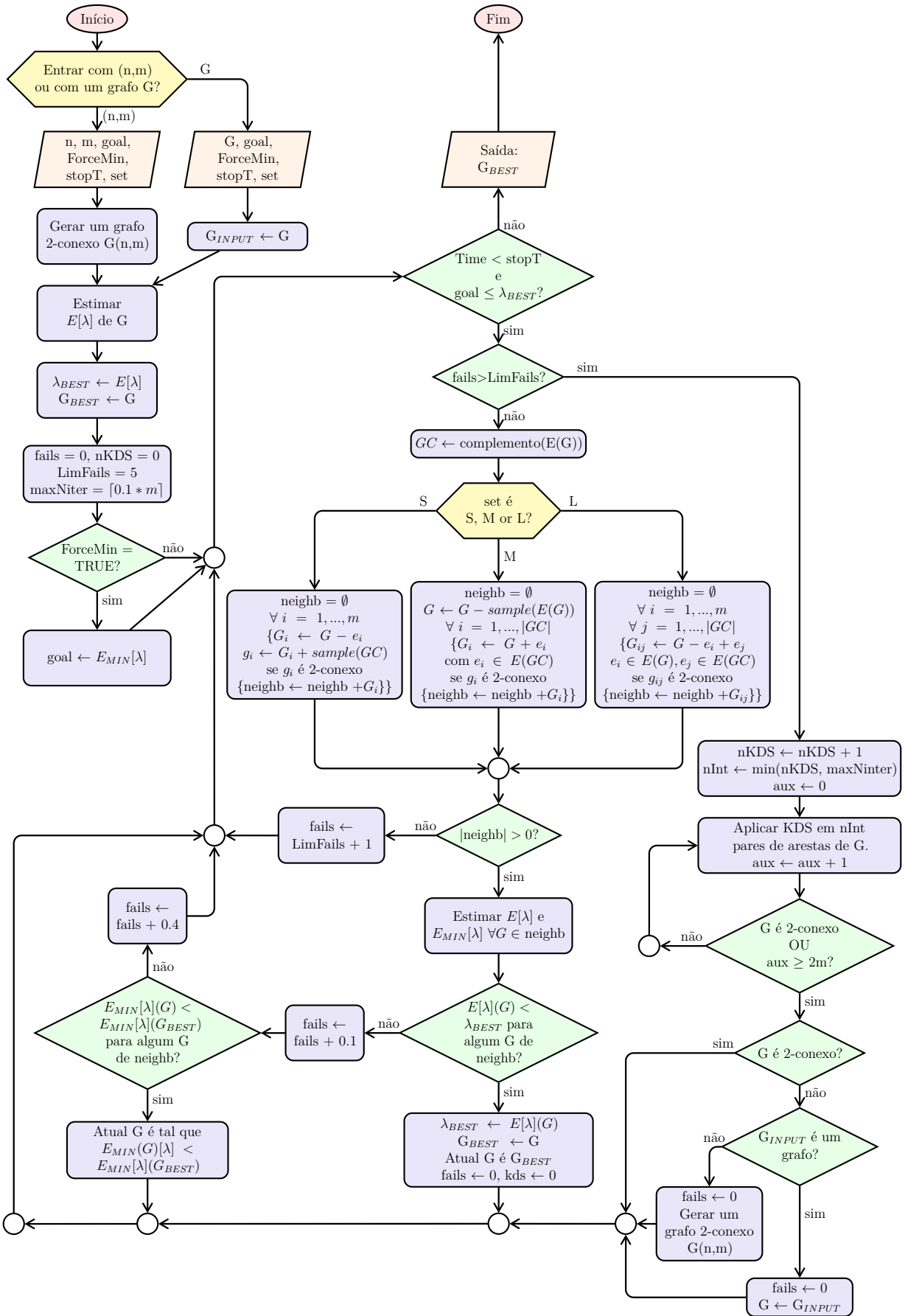


Figura 24 – Heurística da busca local com inspeção completa das vizinhanças.

também denominada *fails* assim como para *Downhill*, para identificar e tomar providência quanto à alta frequência de falhas que atrasem a busca em questão. No caso, a providência tomada também é a mesma do caso *Downhill*, ou seja, é aplicado o método KDS para troca do local da vizinhança.

Por fim, propõe-se a heurística, com algoritmo dado na Figura 25, advinda do conceito de busca em vizinhança variável (o chamado VNS). Heurísticas VNS evitam ficar presas em mínimos locais, aumentando a cobertura da vizinhança de busca quando a solução não está melhorando, e diminuindo a cobertura se for verificado que a vizinhança é propícia a bons resultados (HANSEN; MLADENOVIC, 2014). Essa proposta também possui como entradas os parâmetros (n, m) ou $G, goal, ForceMin, stopT$, os quais possuem as mesmas funções que nas demais heurísticas apresentadas até aqui. O diferencial aqui é o parâmetro de entrada *Ntype*, o qual propicia a ocorrência de duas abordagens de VNS propostas para essa heurística. Sendo assim, quando na entrada *Ntype* = Edges, a vizinhança gerada fica dada pela retirada de *vns* (com *vns* inicialmente igual a 1) arestas aleatórias do grafo G , e a inserção de *vns* arestas também aleatórias do complementar de G . Esse procedimento é repetido um número *NhS* (estipulado aqui igual a m) de vezes, e apenas os grafos 2-conexos resultantes são considerados, gerando assim uma vizinhança com até *NhS* grafos.

Se, para esse caso de *Ntype* = Edges, o contador de falhas *fails* ultrapassar um valor limite pré-estabelecido, o *LimVNS* (que nesse caso vale 2), o *vns* é então incrementado em uma unidade (com *vns* limitado a *MaxKDS*), de modo a gerar uma vizinhança que abranja grafos mais diversos pela retirada e inserção de uma maior quantidade de arestas. Encontra-se aí o ponto em que pode-se denominar essa heurística de VNS, pois o modo de criação das vizinhanças é modificado, caso observe-se falta de evolução na busca dos resultados. Lembrando que o contador *fails* na presente heurística tem o mesmo papel que nas heurísticas anteriores, que é o de identificar que não estão sendo gerados grafos 2-conexos, e de apontar que as vizinhanças de grafos geradas não estão contendo topologias de redes que minimizem $E[\lambda]$, diante da solução atual. Mais especialmente, se *fails* ultrapassar o parâmetro interno *LimFails* (que nesse caso vale 5), é aplicado o método KDS (tratado anteriormente) ao grafo atual, para buscar sair de “buracos” locais e gerar vizinhanças mais prósperas em termos de solução. Observa-se que *LimFails* deve ser maior que *LimVNS*.

Se, por outro lado, tivermos como entrada *Ntype* = KDS, a vizinhança será montada de outra forma. Nesse caso, é aplicado o método KDS em *vns* pares de arestas do grafo G (também começando com *vns* = 1), um número *NhS* de vezes, gerando assim uma vizinhança com até *NhS* grafos, uma vez que apenas os grafos 2-conexos são considerados. Igualmente, assim que o parâmetro *fails* for sendo incrementado, por motivos de não encontrar redes 2-conexas ou de a solução não avançar, e ultrapassar o parâmetro *LimVNS*,

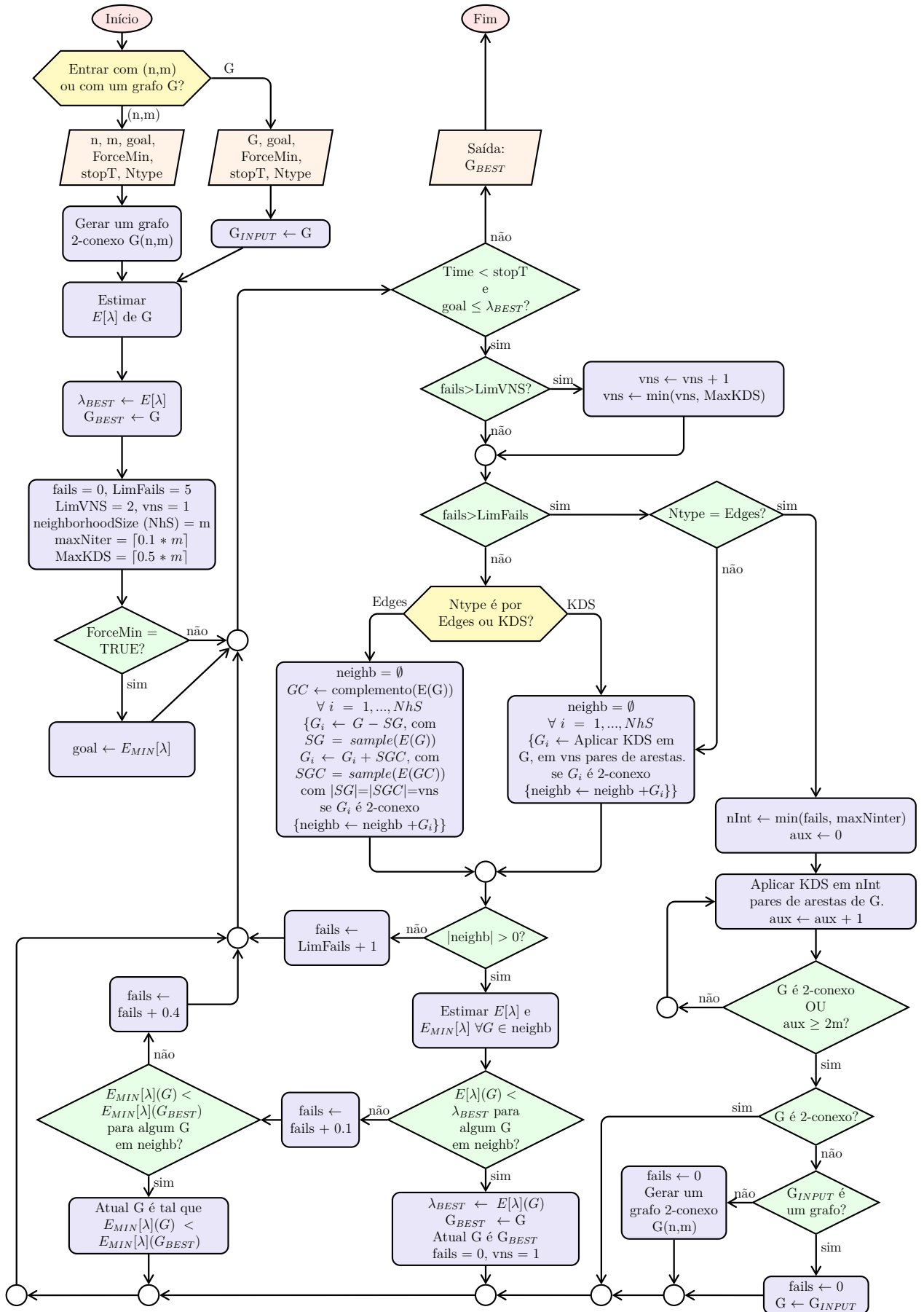


Figura 25 – Heurística baseada na busca em vizinhança variável.

então o parâmetro vns é incrementado (e também limitado a $MaxKDS$), com a esperança de que, com essa maior variabilidade aplicada ao grafo atual, melhores soluções possam ser encontradas. Nesse caso, mesmo que $fails$ ultrapasse $LimFails$, não é aplicado o método KDS para mudança do local da vizinhança, uma vez que esse é o método usado para gerar a própria vizinhança.

Na seção seguinte, são apresentadas aplicações das heurísticas aqui apresentadas, implementadas em linguagem computacional R.

3.3 Aplicações das heurísticas em experimentos computacionais

Na presente seção, as heurísticas propostas na seção anterior são aplicadas em duas abordagens. Primeiro, na Seção 3.3.1 um conjunto de redes reais é testado e topologias alternativas que exigem menor λ são encontradas, onde na Seção 3.3.2 a confiabilidade é comparada entre essas redes reais e suas redes alternativas encontradas. Segundo, na Seção 3.3.3 uma aplicação com rede de entrada aleatória é apresentada.

Para as estimativas de λ , nas heurísticas, são utilizadas as Equações 3.3 e 3.4, logo as redes que são consideradas nos testes das duas abordagens devem atender às hipóteses consideradas na obtenção dessas equações que, no presente caso, são as restrições topológicas dadas na Seção 2.1.1. Caso se queira aplicar as heurísticas para buscar redes que atendam a outro conjunto de hipóteses topológicas, diferentes das que foram usadas para gerar tais equações, os passos de 1 a 5 da Figura 18 devem ser refeitos, tomando então uma amostra S de grafos que considere e atenda o novo conjunto de hipóteses pretendidas, lembrando o fato de que são necessários aproximadamente 12 minutos de tempo computacional para obter as duas equações estimadoras de λ .

As aplicações das heurísticas são todas implementadas no programa R, utilizando o pacote *igraph*, o qual possui muitas funções para a manipulação de grafos e também para o cálculo de invariantes de grafos, o que facilita enormemente a implementação.

3.3.1 Encontrando redes alternativas a redes reais

Para avaliar as potenciais economias que boas redes ópticas alternativas podem gerar no valor de λ , todas as heurísticas propostas são implementadas no programa R e então aplicadas a um conjunto de 12 redes reais, das 15 constantes na Tabela 1, as quais satisfazem as restrições topológicas consideradas para fornecer as Equações 3.3 e 3.4. Definindo “topologias similares” como aquelas que atendam às mesmas restrições topológicas e difiram pelo arranjo de arestas, o objetivo desta aplicação é encontrar topologias similares às topologias reais, mas com melhorias consideráveis com respeito ao requerimento de comprimentos de onda, ou seja, com menores valores de λ .

Executando novamente o método ILP de cálculo exato, dado por Cousineau *et al.* (2015), e aplicando as Equações 3.3 e 3.4 às 12 redes reais consideradas, obtém-se os valores de λ , de $E[\lambda]$ e de $E_{min}[\lambda]$ para cada rede. A Tabela 7 exhibe as redes consideradas com seus valores de n e m , e seus respectivos resultados, onde é possível avaliar o quão longe uma rede está do valor esperado de λ e do seu menor valor teórico possível.

Vale ressaltar que existem redes reais com valores de $E[\lambda]$ inferiores à sua solução exata de λ , o que significa que essas redes não permitem uma alocação de caminho óptico tão boa quanto uma topologia similar média (que são topologias que atendam às mesmas hipóteses topológicas das redes reais em questão, mas que possuam o valor exato de λ próximo na estimativa média $E[\lambda]$). Por exemplo, encontramos redes que funcionam melhor (por exemplo, CANARIE), pior (por exemplo, MZIMA) e igual (por exemplo, NSFNET) a suas respectivas topologias similares médias.

A coluna E_{min} (dada pela Equação 3.4) contém os valores estimados mais baixos de λ que uma boa topologia alternativa, que atende às mesmas restrições topológicas, teoricamente poderia assumir. Comparando λ e $E_{min}[\lambda]$, tem-se a coluna “Economia Potencial”, dada por $100 \times (\lambda - E_{min})/\lambda$, a qual apresenta a porcentagem de economia que teoricamente poderia ser obtida se uma topologia mais adequada fosse escolhida. Esta última coluna da Tabela 7 revela que, para este conjunto particular de topologias reais, a economia teórica varia de 8% a 41%.

Tabela 7 – Comparação entre λ , $E[\lambda]$ e $E_{min}[\lambda]$.

Redes Reais	n	m	λ ILP	E	E_{min}	Economia Potencial
VBNS	12	17	19	18	14	26%
CESNET	12	19	15	14	11	27%
NSFNET	14	21	13	13	12	8%
AUSTRIA	15	22	18	17	14	22%
MZIMA	15	19	30	27	20	33%
ARNES	17	20	38	46	26	32%
GERMANY	17	26	24	25	18	25%
SPAIN	17	28	22	21	16	27%
MEMOREX	19	24	48	58	29	40%
CANARIE	19	26	44	47	26	41%
EON	19	37	17	21	15	12%
ARPANET	20	32	33	33	21	36%

A partir da implementação das heurísticas, estas foram executadas, e a maioria delas demonstrou convergência para uma boa solução em apenas alguns segundos, para os tamanhos de rede aqui tratados, isso em máquinas comuns e acessíveis na atualidade².

²Este tempo de computação refere-se a um notebook Intel Core i3 CPU M 350 com 6 GB DDR3 de RAM.

Devido a essa rápida convergência, as heurísticas são então aplicadas 30 vezes em cada rede real para se verificar a consistência na obtenção da solução, com um *stopT* de 60 segundos para cada execução, e com todos os parâmetros internos dados pelos valores padrão (*default*) já definidos nos fluxogramas das heurísticas. Quanto aos parâmetros de entrada (os *inputs*), para todos os experimentos foi considerado *ForceMin* = FALSE (logo, *goal* não era calculado) e assim, apenas o tempo *stopT* era o critério de parada das heurísticas. Agora, particularmente a cada heurística, foi considerado o seguinte:

- **Busca Aleatória:** foi considerado o n e o m da rede real em questão, já que nesse caso não há a possibilidade de se entrar com uma rede;
- **Downhill (Guloso):** a rede G de entrada é cada rede real dada (sem entrada de valores de n e m), e as duas formas de geração de vizinhança foram consideradas, ou seja, contemplou-se as duas instâncias de *sorted* (TRUE e FALSE);
- **Busca Local:** a rede G de entrada é cada rede real dada (sem entrada de valores de n e m), e as três formas de geração de vizinhança foram consideradas, ou seja, contemplou-se as três instâncias de *set* (S, M e L);
- **VNS:** a rede G de entrada é cada rede real dada (sem entrada de valores de n e m), e as duas formas de geração de vizinhança foram consideradas, ou seja, contemplou-se as duas instâncias de *Ntype* (Edges e KDS).

Após 30 aplicações de cada heurística (para uma rede de entrada ou para uma instância de n e m), o primeiro melhor resultado na ordem de execução dentre os 30, ou seja, a aplicação que primeiro forneceu uma rede com o menor $E[\lambda]$ dentre os 30, é considerada a aplicação que fornece a rede solução da heurística, e soluções que aparecem na sequência com o mesmo valor mínimo de $E[\lambda]$ da primeira solução são desconsideradas, mesmo que as soluções posteriores tenham sido obtidas com um tempo de execução mais curto.

Os resultados desses esforços são mostrados no Apêndice F, onde pode-se observar os resultados da aplicação de cada heurística executada para cada tipo de vizinhança possível. Tem-se então, nesse apêndice, a exibição dos valores de: (i) $A = \lambda$ exato da rede original, (ii) $B = E[\lambda]$ de G_S (onde G_S é o grafo solução fornecido pela heurística), (iii) $C = \lambda$ exato de G_S , (vi) Diferença entre estimativa e valor obtido: $C - B$ (abs), (v) Diferença entre estimativa e valor obtido: $C - B$ (em %), (iv) Melhora Total: $A - C$ (abs), (vii) Melhora Total: $A - C$ (em %), obtidos para cada rede real testada.

A Tabela 8 compila os resultados apresentados no Apêndice F, focando no λ exato obtido para o grafo solução G_S fornecido por cada heurística, para cada vizinhança de busca possível. Nessa tabela, as células preenchidas com a cor rosa são aquelas que guardam os menores valores encontrados por rede, dentre todas heurísticas. Dado que as redes estão ordenadas por sua ordem n (mesma ordenação da Tabela 7), passando pelas linhas da

tabela nota-se primeiramente que, quanto maior a ordem da rede, a quantidade de células coloridas por rede tende a diminuir, ou seja, menor é a quantidade de soluções igualmente menores.

Redes Reais	λ	Busca Aleatória	Downhill		Busca Local			VNS	
			<i>sorted =</i> FALSE	<i>sorted =</i> TRUE	<i>set =</i> S	<i>set =</i> M	<i>set =</i> G	<i>Ntype =</i> Edges	<i>Ntype =</i> KDS
VBNS	19	9	9	9	9	9	9	9	9
CESNET	15	7	8	8	7	7	7	8	8
NSFNET	13	10	10	13	10	10	10	10	10
AUSTRIA	18	12	12	13	12	11	12	12	11
MZIMA	30	15	16	17	16	16	15	15	15
ARNES	38	23	22	24	22	22	22	23	22
GERMANY	24	13	13	14	14	13	14	13	13
SPAIN	22	12	13	14	11	11	12	12	11
MEMOREX	48	24	22	25	21	21	22	22	21
CANARIE	44	19	18	21	18	19	19	19	18
EON	17	11	11	16	10	10	13	10	11
ARPANET	33	16	16	33	15	15	20	15	15

Tabela 8 – Valores de λ das redes reais de entrada versus os valores de λ das redes de saída (G_S) das heurísticas, para cada vizinhança possível. As células preenchidas de rosa são aquelas que guardam o valor mínimo encontrado por rede dentre todas heurísticas.

Atribui-se a esse ocorrido o fato de que, com o aumento da ordem n da rede, aumentam também (e de modo exponencial) a quantidade de redes a serem varridas, provavelmente dificultando aí a busca daquelas redes que possam requerer um λ menor. Essa afirmação ganha suporte quando se observa a coluna dos resultados da heurística **busca aleatória**, onde a busca é randômica, e resultados de λ tão menores quanto o de outras heurísticas deixam de ser encontrados a medida que observa-se redes maiores. Sendo assim, pode-se ver que a busca aleatória, como era de se esperar, não se mostra consistente com o aumento do tamanho da rede, mesmo utilizando-se de uma função substituta otimizada na estimativa de λ .

Analisando agora as duas execuções da heurística *Downhill*, obteve-se um melhor desempenho para o caso de *sorted = FALSE* do que no caso *sorted = TRUE*. Nesse caso a aleatoriedade (numa forma mais controlada do que no caso de **busca aleatória**) trabalhou a favor, pois a inserção da aresta no caso *sorted = FALSE* é aleatória, ao contrário do outro caso, cuja inserção de arestas se dá a partir de uma ordem pré-definida. A aleatoriedade foi capaz aí de fornecer soluções de modo mais independente do tamanho da rede, do que a heurística **busca aleatória**.

De modo complementar, vê-se que mesmo que a heurística *Downhill* tome o primeiro melhor resultado sem verificar toda a vizinhança, existe a limitação de tamanho da vizinhança a ser varrida, dada como parâmetro interno, que no caso é m (em “*pos > m?*”

na heurística), o mesmo m do número de arestas. Então vê-se que no caso de $sorted = \text{TRUE}$, onde a inserção de arestas ocorre de modo ordenado, as arestas inseridas ficam limitadas sempre às m primeiras do grafo complementar, limitando assim a busca de grafos de modo mais diverso. Para contornar essa situação, o tamanho da vizinhança possível de ser varrida deveria se incrementado.

A heurística de *Busca Local* foi executada em suas três possibilidades de vizinhança, as quais são inteiramente varridas nas diversas iterações. Observa-se nos resultados da Tabela 8 que, para $set = S$ e M , obteve-se os melhores resultados, com pequena vantagem para $set = M$, com um melhor resultado a mais que $set = S$. Apenas para duas redes reais, a opção $set = M$ não encontrou a melhor solução comparativamente às demais heurísticas. Lembrando que $set = S$ cria novas redes do modo como em $sorted = \text{FALSE}$ na heurística *Downhill*, e que $set = M$ cria novas redes de um modo parecido com o de $sorted = \text{TRUE}$ na heurística *Downhill*³, aqui, ao contrário do que ocorreu na simulação por *Downhill*, o segundo modo se destacou pois o tamanho da vizinhança não estava limitado, de forma que foi possível verificar então a eficiência desse segundo modo.

A opção $set = G$ deixou de entregar bons resultados a partir do momento em que a rede real de entrada crescia. Atribui-se isso ao fato de que, mesmo esse caso sendo uma extensão do método usado em $set = M$, aqui a vizinhança gerada é possivelmente demasiadamente grande, de modo que com redes de entrada cada vez maiores, gera-se vizinhanças proporcionalmente maiores (as quais têm tamanho dado em função de n e m), e gasta-se assim mais tempo em localidades de vizinhança com talvez pouco potencial. A solução então aí seria incrementar o tempo necessário para obter melhores soluções, caso deseje-se realmente utilizar essa via na busca de redes com um menor requerimento de λ .

Por fim, a heurística VNS foi aplicada em suas duas instâncias, onde obteve melhores resultados em $Ntype = \text{KDS}$ do que em $Ntype = \text{Edges}$, mesmo que em ambos os casos o tamanho da rede não tenha influenciado muito no desempenho das heurísticas, fato este possivelmente influenciado pelo próprio conceito de VNS, mais avançado que os demais conceitos de heurísticas usados até aqui. Já a vantagem da abordagem $Ntype = \text{KDS}$ deve-se possivelmente ao fato de que o método KDS propicia uma diversificação maior da vizinhança gerada, do que a retirada e inserção de um conjunto de arestas feito quando $Ntype = \text{Edges}$.

Feita toda esta análise, pode-se dizer que houveram duas melhores instâncias, empatadas em número de melhores resultados, que foram a Busca Local com $set = M$, e o VNS com $Ntype = \text{KDS}$, e elegemos a segunda heurística, pela sua maior sofisticação, para construir uma análise mais aprofundada de seus resultados, os quais são apresentados na Tabela 9. A título de simplificação, chamemos essa heurística VNS com $Ntype = \text{KDS}$

³Só diferindo do fato de que apenas uma aresta aleatória é retirada, e cada aresta do grafo complementar inserida gera uma rede de teste.

de VNS-KDS. Acredita-se que a heurística VNS-KDS obteve maior êxito dada a maior possibilidade de criação de vizinhanças internamente mais diversificadas (do que o VNS com $Ntype = Edges$), isso graças ao método KDS.

























As primeiras duas colunas da Tabela 9 contém o nome e uma exibição gráfica das topologias de redes reais de entrada usadas nas heurísticas. As colunas “ λ ILP” e “Economia potencial” são as mesmas já exibidas na Tabela 7, e são apresentadas também aqui a título de facilitar a comparação de resultados. Na coluna “Topologias resultantes”, as melhores soluções da heurística VNS-KDS com melhores valores de λ são exibidas, as quais diferem da topologia real de entrada apenas por um rearranjo de enlaces, uma vez que a troca das posições das arestas é de fato a ação usada pela heurística para buscar melhores redes, mantendo assim na rede de saída o mesmo n e m da rede real de entrada. No Apêndice G encontra-se a lista de arestas das redes reais de entrada e das redes soluções exibidas graficamente na Tabela 9.

A coluna “Menor $E[\lambda]$ encontrado” mostra os valores de λ esperado para as melhores soluções, e a próxima coluna “ λ ILP da melhor solução” exhibe o valor de λ exato para essas topologias. A coluna “Economia real” é dada por $100 \times (\lambda_E - \lambda_S)/\lambda_E$, onde λ_E é o λ exato da rede real de entrada e λ_S é o da rede alternativa de saída. Essa coluna traz a proporção da verdadeira economia de λ em todo o processo, comparando o valor de λ da rede inicial com o da nova rede. A economia de λ obtida com a nova topologia é considerável, variando de 23% (em NSFNET) até 59% (em CANARIE), e superando as economias potenciais estimadas anteriormente. Essa superação é possível, ou mesmo previsível, porque a heurística busca a melhor configuração de topologia passando por várias delas, melhores que a rede de entrada inclusive, o que vai alterando os parâmetros de topologia, diminuindo assim a estimativa de $E_{min}[\lambda]$ a cada melhor rede encontrada. O $E_{min}[\lambda]$ inicial refere-se aos parâmetros da topologia de rede real, e a solução escolhida tem novas e melhores estimativas. Por outro lado, $E[\lambda]$ e λ são valores muito próximos nas topologias resultantes, o que garante que a amostra aleatória inicial S usada na obtenção das equações estimadoras foi representativa o suficiente, tanto qualitativa quanto quantitativamente.

As últimas três colunas da Tabela 9 tratam de questões de desempenho heurístico. A primeira dessas três colunas, a “Tempo gasto p/ obter a solução”, apresenta o intervalo de tempo gasto⁴ para chegar à melhor solução nas 30 aplicações da heurística VNS-KDS. Por exemplo, para a rede NSFNET, nas 30 rodadas de um minuto da heurística, a melhor solução em cada rodada ocorreu entre 0,64 segundos e 55,67 segundos. Nessa mesma coluna, o tempo entre parênteses é o tempo gasto na rodada que gerou a topologia apresentada como solução final. Ainda no caso da NSFNET, a solução apresentada ocorreu, em uma

⁴Este tempo de computação também refere-se a um notebook Intel Core i3 CPU M 350 com 6 GB DDR3 de RAM.

Tabela 9 – Detalhamento dos resultados das redes reais para a heurística VNS-KDS.

Redes Reais		Número de topologias similares	λ ILP	Economia potencial	RESULTADOS PARA A HEURÍSTICA VNS-KDS						
					Topologias resultantes	Menor $E[\lambda]$ encontrado	λ ILP da melhor solução	Economia real	Tempo gasto p/ obter a solução	Topologias visitadas para obter a solução	% dos testes com $E[\lambda] \leq \lambda$
VBNS		331005 (< 1 s)	19	26%		8	9	53%	0,39s - 36,79s (36,79s)	10800 - 11190 (11146 ~ 3,37%)	100% (100%)
CESNET		7274750 (7 s)	15	27%		7	8	47%	0,43s - 36,79s (1,1s)	10800 - 11440 (11100 ~ 0,15%)	100% (100%)
NSFNET		161191179 (190 s)	13	8%		9	10	23%	0,64s - 55,67s (7,74s)	9999 - 10389 (10347 ~ 0,006%)	100% (100%)
AUSTRIA		629554890 (1080 s)	18	22%		10	11	39%	0,65s - 36,84s (31,62s)	8704 - 9453 (9347 ~ 0,0015%)	100% (13,3%)
MZIMA		549306 (16 s)	30	33%		14	15	50%	1,18s - 47,94s (15,5s)	6562 - 7550 (7550 ~ 1,37%)	100% (80%)
ARNES		72639 (68 s)	38	32%		20	22	42%	1,34s - 56,25s (2,83s)	3651 - 5947 (5864 ~ 8,07%)	96,7% (50%)
GERMANY		526560187393 (747713 s)	24	25%		12	13	46%	2,09s - 56,66s (10,95s)	7590 - 8030 (8014 ~ 1,5 × 10⁻⁶%)	100% (43,3%)
SPAIN		<i>Time out*</i>	22	27%		10	11	50%	3,60s - 47,33s (6,74s)	6892 - 7631 (7485)	100% (23,3%)
MEMOREX		198398357 (35063 s)	48	40%		20	21	56%	0,80s - 58,17s (9,84s)	4860 - 6743 (6431 ~ 0,003%)	33,3% (3,3%)
CANARIE		62314889998 (851271 s)	44	41%		17	18	59%	1,11s - 38,86s (4,53s)	5659 - 6926 (6606 ~ 1,1 × 10⁻⁵%)	30% (3,3%)
EON		<i>Time out*</i>	17	12%		9	11	35%	2,10s - 38,74s (10,28s)	6569 - 6967 (6880)	100% (6,7%)
ARPANET		<i>Time out*</i>	33	36%		13	15	55%	1,85s - 55,70s (4,74s)	6362 - 6582 (6519)	100% (13,3%)

* Mais de um mês de cálculo contínuo não foi suficiente para completar a contagem de topologias similares.

das 30 rodadas, no tempo de 7,74 segundos dentre os 60 segundos da execução. O pequeno tempo gasto para alcançar as boas soluções, mesmo para redes maiores (como pode ser verificado na referida tabela), reforça a eficiência do método. O cálculo exato de λ via ILP ocorre apenas no momento de verificação da topologia resultante.

A penúltima coluna da Tabela 9, a “Topologias visitadas para obter a solução”, informa o número de topologias visitadas por cada uma das 30 rodadas da heurística. Citando novamente a NSFNET, foram visitadas entre 9999 e 10389 topologias em cada uma das 30 rodadas dessa topologia. Além disso, também é apresentado nesta coluna o número total de topologias visitadas na rodada que fornece a topologia resultante apresentada (também entre parênteses), onde, no caso da NSFNET, a rodada que forneceu a topologia solução varreu 10347 topologias. Nesta coluna, vê-se que o número de topologias visitadas em cada rodada diminui quando o tamanho da topologia aumenta. Essa diminuição ocorre porque uma vizinhança com topologias maiores, com mais vértices e arestas, provavelmente demora mais para ser montada e toma mais tempo no cálculo de seus invariantes.

Os valores da terceira coluna da Tabela 9, a “Número de topologias similares”, exibem a quantidade total de topologias similares 2-conexas de cada topologia de rede real, a qual é calculada no programa *Nauty*. O tempo em segundos ao lado do número de topologias similares é o tempo gasto para computar tal número. Observando que tais tempos crescem exponencialmente, constata-se também aí a importância de se ter um método eficiente de busca de redes que atendam a critérios pré-estabelecidos, uma vez que com o aumento da ordem da rede, torna-se inviável até saber quais são todas as redes possíveis, e quiçá testar todas.

É interessante comparar, porém, os valores dessa coluna de topologias similares com os valores entre parênteses na coluna “Topologias visitadas para obter a solução”. A comparação entre o número total de topologias visitadas pela heurística VNS-KDS com o número de topologias similares também é apresentada, como uma porcentagem, na coluna de topologias visitadas. Essa porcentagem mostra que, em todos os casos, apenas uma pequena proporção de possíveis topologias é visitada, para encontrar uma solução bastante razoável. Essas proporções e os pequenos tempos gastos para encontrar uma boa solução nos indicam que os estimadores $E[\lambda]$ e $E_{min}[\lambda]$ conduzem bem a heurística VNS-KDS no caminho para encontrar uma boa solução, no que tange o requerimento de comprimentos de onda.

Finalmente, a última coluna “% dos testes com $E[\lambda] \leq \lambda$ ” mostra a proporção das 30 rodadas da heurística que tiveram seu λ final valendo pelo menos $E[\lambda]$. Considera-se essa comparação aqui porque, em todos os casos, o valor de λ exato da topologia solução final variou de uma a duas unidades acima do valor estimado $E[\lambda]$. Considera-se o alto valor dessa frequência como algo positivo, pois infere-se com isso que $E[\lambda]$ foi bem minimizado, de modo que o cálculo de λ exato, sempre acaba sendo um valor um pouco maior. Então

busca-se, dessa forma, verificar a ocorrência desse fato nas 30 rodadas de cada rede, e constata-se uma alta porcentagem das rodadas com $E[\lambda] \leq \lambda$ em boa parte das redes executadas. Para as redes MEMOREX e CANARIE o resultado dessa coluna não foi tão bom, e não temos clareza do motivo para isso, que pode ser atribuído por exemplo ao fato de tais redes não estarem eventualmente inseridas em boas vizinhanças de grafos.

Nessa última coluna também é apresentado, entre parênteses, a porcentagem das rodadas que forneceram o menor valor de $E[\lambda]$, ou seja, o valor de $E[\lambda]$ que indica a rede solução. Principalmente para topologias maiores, essa segunda proporção é baixa, provavelmente devido à menor proporção de topologias visitadas nesses casos. Então, para essas redes maiores, pode ser útil aumentar o número de rodadas ou o tempo de cada execução, ou uma combinação das duas coisas, mas nada que mude a ordem de grandeza do tempo necessário para obtenção de uma solução ainda melhor.

3.3.2 Comparação da confiabilidade das redes de entrada e de saída

Uma pergunta natural que surge a partir dos resultados mostrados até aqui, é se redes com menor requerimento de comprimentos de onda possuem outras características de interesse também melhoradas, comparativamente às suas redes similares, além de apenas uma maior disposição do espectro. Quando se trata de criar uma rede óptica, várias questões são de interesse dos projetistas de redes, algumas já ditas aqui, como o CAPEX e a probabilidade de bloqueio. Escolhe-se tratar aqui, porém, da confiabilidade nas redes de saída, comparativamente às redes de entrada. A confiabilidade de uma rede, dada pela probabilidade da rede permanecer conexa após a falha de um conjunto de enlaces (PAVAN *et al.*, 2015), é um quesito de grande interesse em redes, pois redes mais confiáveis são, de certo modo, redes mais resilientes, uma vez que são mais capazes de seguir atendendo uma dada demanda de tráfego, mesmo com danos nas estruturas de certos enlaces.

Assumindo que enlaces de uma rede G podem falhar de modo independente e com uma probabilidade $\rho \in (0, 1)$, Kel'mans (1967) *apud* Pavan *et al.* (2015) diz que a confiabilidade ($R_G(\rho)$) dessa rede é dada pela probabilidade:

$$R_G(\rho) = 1 - \sum_{i=k'(G)}^m S_i \rho^i (1 - \rho)^{m-i}, \quad (3.5)$$

onde $k'(G)$ é a conectividade de arestas de G , e S_i é o número de conjuntos de corte de arestas de tamanho i , definidos na Seção 1.3.

Mesmo que, de um ponto de vista prático, a probabilidade ρ de falha do enlace geralmente não seja idêntica para cada enlace da rede, e também que tais falhas nem sempre sejam independentes, essas simplificações permitem explorar mais facilmente a confiabilidade. Assim como em Pavan *et al.* (2015), para o cálculo de $R_G(\rho)$ usa-se aqui o valor escolhido de $\rho = 0,05$, uma vez que a probabilidade de ruptura de enlaces em

redes ópticas é usualmente muito baixa. Observa-se que maximizar $R_G(\rho)$ implica em maximizar $k'(G)$ e minimizar S_i , o que na literatura é conhecido como o problema de $\max k'(G) \& \min S_i$ (PAVAN *et al.*, 2015).

O cálculo de $R_G(\rho)$ é *NP-hard*, fato esse devido à complexidade computacional do cálculo de S_i para todo i (PROVAN; BALL, 1983). Diante disso, opta-se por restringir o cálculo de S_i para $i = 2, 3, 4$, assim como em Pavan *et al.* (2015), onde o i inicia em dois por causa da 2-conexidade das redes tratadas em questão. O cálculo de S_i é feito no programa R, também usando o pacote *igraph*, e seu resultado é dado na Tabela 10. Observa-se nessa tabela que, em todos os casos, os valores de S_i não aumentaram nas redes de saída (RS) comparativamente às redes reais (RR) de entrada, o que já indica que a confiabilidade, para as redes de saída em questão, certamente não será menor do que a das redes de entrada.

	REDE	S_2	S_3	S_4	REDE	S_2	S_3	S_4
Rede Real (RR)	VBNS	7	123	940	GERMANY	10	255	3033
Rede de Saída (RS)		4	70	563		7	177	2120
Diferença (RR-RS)		3	53	377		3	78	913
Rede Real (RR)	CESNET	6	108	901	SPAIN	12	310	3806
Rede de Saída (RS)		6	104	837		10	262	3265
Diferença (RR-RS)		0	4	64		2	48	541
Rede Real (RR)	NSFNET	2	51	596	MEMOREX	23	524	5399
Rede de Saída (RS)		2	49	557		10	252	2977
Diferença (RR-RS)		0	2	39		13	272	2422
Rede Real (RR)	AUSTRIA	14	284	2628	CANARIE	17	434	5054
Rede de Saída (RS)		6	127	1271		10	252	3017
Diferença (RR-RS)		8	157	1357		7	182	2037
Rede Real (RR)	MZIMA	14	276	2266	EON	6	214	3702
Rede de Saída (RS)		9	181	1620		5	179	3105
Diferença (RR-RS)		5	95	646		1	35	597
Rede Real (RR)	ARNES	36	588	4068	ARPANET	0	20	618
Rede de Saída (RS)		14	306	2962		0	16	485
Diferença (RR-RS)		22	282	1106		0	4	133

Tabela 10 – Valores de S_2 , S_3 e S_4 para todas as redes reais em análise.

O resultado do cálculo da confiabilidade $R_G(\rho)$ é então dado na Tabela 11. De fato, verifica-se nessa tabela que todos os valores de confiabilidade das redes de saída são maiores do que os das redes reais de entrada, mostrando uma tendência de que redes com um menor requerimento de comprimentos de onda possam ser mais resistentes a falhas simultâneas num conjunto de arestas, tendência essa que pode vir a ser melhor verificada no futuro.

Rede Real	$R_G(\rho)$ para as Redes Reais (RR)	$R_G(\rho)$ para as Redes de Saída (RS)	Diferença (RR - RS)
VBNS	0,9814	0,9893	0,0079
CESNET	0,9852	0,9856	0,0004
NSFNET	0,9940	0,9942	0,0002
AUSTRIA	0,9675	0,9855	0,0179
MZIMA	0,9636	0,9759	0,0123
ARNES	0,9223	0,9620	0,0396
GERMANY	0,9768	0,9838	0,0070
SPAIN	0,9744	0,9784	0,0040
MEMOREX	0,9470	0,9745	0,0275
CANARIE	0,9607	0,9769	0,0162
EON	0,9886	0,9904	0,0019
ARPANET	0,9985	0,9988	0,0003

Tabela 11 – Valores de $R_G(\rho)$ para todas as redes reais em análise e suas respectivas redes de saída.

3.3.3 Começando por uma topologia aleatória

Em relação ao projeto de redes ópticas, duas situações principais podem ocorrer. No primeiro cenário, a rede cresce com o passar do tempo e enlaces são adicionados gradualmente. Num segundo cenário, para atender a algumas metas de capacidade e desempenho, a rede inteira é projetada e definida de início. A questão natural que segue é qual a diferença entre essas duas abordagens, com relação a um mesmo critério de interesse. Diante disso, a presente seção compara a diferença entre um cenário de expansão de rede e o de projeto de toda a rede desde o início, com relação ao requerimento de comprimentos de onda, buscando analisar como essas duas situações podem se comportar.

Para essa análise, esses dois cenários são testados na rede NSFNET, que possui $n = 14$ e $m = 21$. No primeiro cenário, suponha que se queira inserir um enlace na rede NSFNET, a qual possui

$$[n(n - 1)]/2 - m = [14(14 - 1)]/2 - 21 = 70$$

diferentes posições possíveis para fazer isso. Conseqüentemente, pela inserção de um enlace em NSFNET, 70 topologias diferentes podem ser geradas. No segundo cenário, assume-se que, ao fim, se deseje uma NSFNET com $n = 14$ e $m = 22$ (um enlace a mais que a NSFNET original, para condizer com o cenário 1 em que uma aresta é adicionada) e que, nesse segundo caso, uma rede similar com essas propriedades é criada usando a heurística VNS-KDS que foi proposta na Seção 3.2, porém usando agora a opção de início por meio de uma topologia aleatória, ao contrário do que foi feito na Seção 3.3.1.

Nesta comparação, é possível verificar se vale a pena projetar uma rede com as propriedades que ela já deve ter ao fim, do ponto de vista da minimização de λ : considerando uma mesma demanda de tráfego uniforme, os valores de λ nesses dois cenários são comparados para verificar se existe vantagem em projetar uma rede inteira do início (cenário 2), ou se o projeto de uma rede, enlace por enlace, pode ser suficiente (cenário 1).

Começando com o cenário 1, as 70 topologias possíveis são geradas pela inserção de um enlace a cada par de nós sem enlace da NSFNET, onde cada inserção diferente gera uma topologia diferente. Então, o λ exato é calculado para todas essas topologias usando o mesmo método utilizado até agora (COUSINEAU *et al.*, 2015), obtendo-se valores de λ variando de 11 a 14. Pode-se ver então, com isso, que uma inserção de enlace bem escolhida (Figura 26a) tem o potencial de diminuir em até duas unidades o valor de λ , nesse caso. Por outro lado, uma inserção de enlace mal projetada (Figura 26b) tem o potencial para inclusive adicionar uma unidade indesejada em λ . Relembrando que o valor de λ para a rede NSFNET é 13 (Figura 27a). Vale ressaltar que com um enlace adicional a topologia aumentou seu grau médio, o que não necessariamente implicou em decréscimo do valor de λ .

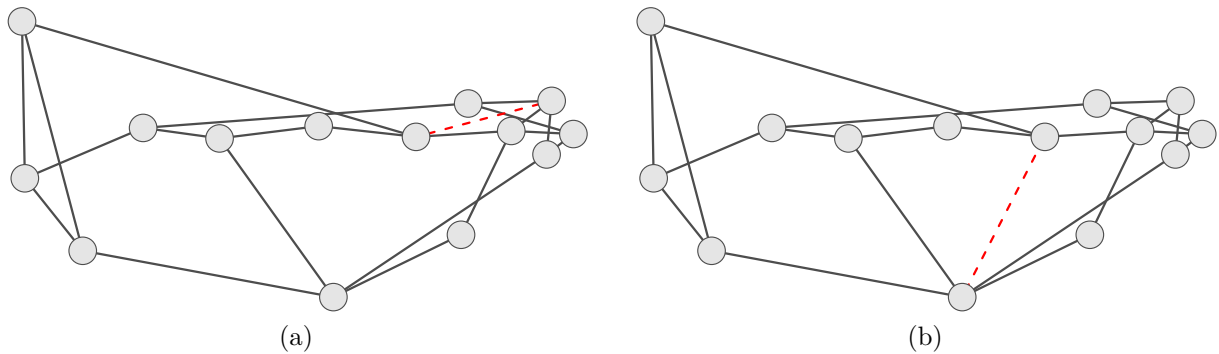


Figura 26 – Cenário 1: Dois casos da NSFNET com um enlace adicionado em vermelho tracejado. Em (a), $\lambda = 11$ (melhor caso) alcançado em 18 das 70 possibilidades, e em (b), $\lambda = 14$ (pior caso) alcançado em 6 das 70 possibilidades.

Agora, no cenário 2, a topologia de NSFNET é construída do início, supondo-se que ela não existe, e apenas com as informações que $n = 14$ e $m = 22$, e mantendo-se as restrições topológicas dadas na Seção 2.1.1. Com essas entradas, a heurística VNS-KDS é executada 30 vezes, com um tempo de 120 segundos para cada rodada. No entanto, a rodada que encontra a melhor solução dentre as 30 faz isso nos seus primeiros segundos, assim como os resultados da dados na Tabela 9. O λ exato é calculado para a primeira topologia dentre as que alcançaram o melhor λ esperado ($E[\lambda]$). O resultado obtido nesse cenário 2 é $\lambda = 9$ (Figura 27b), o qual é 18% melhor que o melhor resultado do cenário 1.

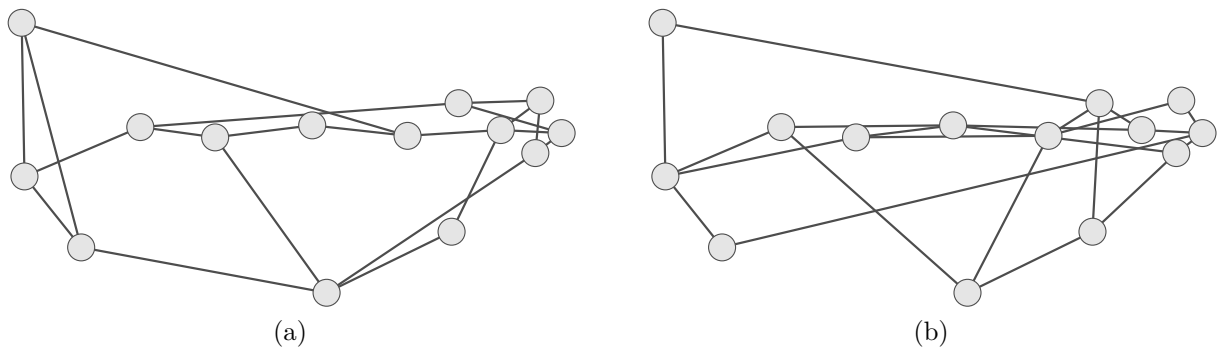


Figura 27 – Em (a), topologia real da NSFNET, com $n = 14$, e $m = 21$, a qual tem $\lambda = 13$. Em (b), Cenário 2: Topologia gerada com $n = 14$, e $m = 22$, a qual possui $\lambda = 9$.

Com essa análise, é mostrado o quão importante é um projeto inicial de uma rede inteira. Chama-se atenção para o fato de que ajustes pontuais em uma rede podem ser mais desastrosos do que se pode imaginar, resultando em desempenho abaixo do potencial.

3.4 Conclusão

Por meio dos invariantes mais bem correlacionados com λ , advindos do Capítulo 2, funções substitutas são geradas e utilizadas em heurísticas, que são implementadas para buscar topologias com menor requerimento de comprimentos de onda. Resultados interessantes são obtidos com redes reais de entrada, chegando a uma economia de até 59% no requerimento de comprimentos de onda nas redes alternativas de saída. E com entrada de rede aleatória mostrou-se, por meio da rede NSFNET, a importância de planejar por completo, e desde o início, a topologia física de uma rede.

4 Conclusões e Trabalhos Futuros

Sistemas inteligentes para o projeto de redes ópticas lidam frequentemente com aspectos conflitantes, como recursos computacionais e de rede (YANG *et al.*, 2010) ou diferentes tipos de fluxos de rede (PRZEWOŹNICZEK *et al.*, 2015). Essas abordagens geralmente são difíceis não apenas por causa de sua complexidade computacional, mas também devido à imensa escala de seu espaço de solução. Esses problemas são extensivamente estudados usando todo tipo de métodos de inteligência artificial, redes neurais e algoritmos genéticos (HANAY; ARAKAWA; MURATA, 2015).

Um tópico amplamente explorado em muitas áreas de sistemas inteligentes é a seleção de atributos (BENNASAR; HICKS; SETCHI, 2015). Até onde se sabe, não há nenhum trabalho anterior na literatura que recorra à seleção de atributos para o projeto de redes ópticas. Alguns trabalhos exploram invariantes da teoria dos grafos para explicar o número de comprimentos de onda, mas não de modo tão abrangente quanto o que é apresentado aqui. No melhor dos esforços, cobre-se a literatura relacionada a esse assunto e agrupa-se uma lista com 315 invariantes topológicos que são mais fáceis de calcular do que o número de comprimentos de onda.

Este estudo explora a influência de cada invariante topológico analisado em redes ópticas, mais especificamente, no requerimento de comprimentos de onda. Isso é feito usando um método de seleção de atributos baseado em informação mútua, similar a Bennasar, Hicks e Setchi (2015) mas, no presente caso, é proposto um estimador que pode ser aplicado a dados discretos ou contínuos, independentemente. Um estimador de informação mútua baseado num estimador de entropia (MONTALVÃO; ATTUX; SILVA, 2014) é adequado para obter estimativas em espaços de alta dimensão, similar ao estimador *Neighborhood Mutual Information* dado por Hu *et al.* (2011). Ambos os estimadores podem ser aplicados a atributos discretos e contínuos e não requerem discretização. Entretanto, o estimador aqui proposto - baseado no método da coincidência - adota uma abordagem mais intuitiva e mais fácil de usar, e não necessita estimar funções de densidade de probabilidade.

A informação mútua ajuda a identificar invariantes essenciais dentre os 315 invariantes analisados. Amostras com 15 redes do mundo real e $2,2 \times 10^6$ topologias aleatórias (que imitam redes reais) são consideradas. Dois tipos de análises são realizadas, uma considerando todas as redes juntas, e outra separando as redes por ordem. Em ambos os casos, é observado que as redes reais e as redes aleatórias em estudo apresentaram comportamento similar com respeito ao requerimento de comprimentos de onda, demonstrando a representatividade da amostra de redes aleatórias considerada.

No geral, os resultados da seleção de atributos aplicada destacam a importância de

considerar variáveis relacionadas a congestionamento de arestas, conectividade, transmissão, distância e grau de vértice, para projetar topologias de redes ópticas com baixo requerimento de comprimentos de onda. Para cada uma dessas categorias, um conjunto de invariantes mais relevantes é apresentado, o qual pode ser usado como base de conhecimento no projeto de redes ópticas. Invariantes associados com a conectividade e com o grau de vértice demonstram maior importância na análise separada por ordem de rede. Todavia, mesmo que o conjunto de invariantes aqui trabalhado seja tão diversificado e extenso quanto possível, novos tipos de invariantes podem emergir, os quais podem também vir a ser candidatos relevantes para explicar o número de comprimentos de onda.

Comparando resultados da literatura com resultados do presente trabalho, alguns achados são reforçados, como a importância dos invariantes: grau médio, conectividade algébrica e distância média. Por outro lado, não se confirma a importância dos seguintes invariantes exaltados na literatura: densidade de arestas, variância do grau e número de árvores geradoras.

Vale ressaltar a significância dos invariantes derivados do *edge betweenness* no projeto de redes ópticas, que estão dentre os que demonstraram melhor performance, além de desempenho mais consistente em toda a investigação de invariantes de tempo polinomial que melhor correspondem ao número de comprimentos de onda. Essa proeminência explica por que os algoritmos de roteamento em [Cousineau et al. \(2015\)](#), baseados no *edge betweenness*, tem uma eficiência excelente ao otimizar o congestionamento. Portanto, reforçamos que trabalhos futuros considerem o uso de *edge betweenness* em algoritmos de roteamento, sempre que for de interesse evitar o congestionamento.

Em seguida, são desenvolvidas heurísticas para projeto e expansão de redes ópticas, que fornecem topologias com um menor requerimento de comprimentos de onda. Essas heurísticas usam equações obtidas por uma análise de regressão apropriada, que se baseia em invariantes selecionados destacados no presente estudo. As equações de regressão são usadas como funções substitutas para estimar o número de comprimentos de onda num modelo de otimização (em inglês, um modelo tipo *surrogate-based optimization*).

Com isso, os invariantes *edge betweenness* máximo e a conectividade algébrica, juntamente com a n e m , são usados como variáveis explicativas em regressões estatísticas de Poisson e quantílica para estimar $E[\lambda]$ (valor esperado) e $E_{min}[\lambda]$ (valor mínimo esperado) de λ , respectivamente. Baseando-se no fato da velocidade da estimação de λ ser bem mais vantajosa do que seu cálculo exato, as heurísticas propostas fornecem, com eficiência e em alguns minutos, topologias 2-conexas com um λ o menor possível, dada uma rede de entrada ou ainda dado somente os valores de n e m desejados para a rede.

As heurísticas são testadas em um conjunto de 12 redes reais, e a heurística que proporciona aqui os melhores desempenhos é uma do tipo VNS, chamada aqui de VNS-KDS, onde as topologias similares de saída fornecem uma economia real de λ de 23% a

59%. É comparada ainda a confiabilidade das redes reais de entrada e das redes de saída, onde verifica-se uma maior confiabilidade dentre as redes de saída, comparativamente às 12 redes reais de entrada.

Em seguida, é testada a diferença entre adicionar um enlace a uma rede real, a NSFNET no caso, versus reprojeta-la do início. Os resultados mostram que um projeto de rede refeito do início, pela heurística VNS-KDS, provê um menor requerimento de λ do que inserir um enlace onde quer que seja inserido na NSFNET.

Destaca-se o fato de que toda a metodologia considerada no presente trabalho, desde a obtenção de invariantes que melhor representem o número de comprimentos de onda até a geração de topologias de redes de saída que minimizem o requerimento de comprimentos de onda, pode ser replicada, com as devidas adequações, para qualquer parâmetro de interesse de qualquer tipo de rede, também difícil de calcular, como é o número de comprimentos de onda.

Com isso, como futuras direções de pesquisa, considera-se interessante produzir outros *rankings* de invariantes para explicar outros parâmetros de redes ópticas também difíceis de calcular, como a confiabilidade, a resiliência, a probabilidade de bloqueio, e o CAPEX. Uma vez que vários parâmetros importantes sejam adequadamente explicados por invariantes de grafo de baixo custo computacional, todos esses resultados poderiam ser combinados em um único sistema para projetar topologias de redes ópticas. Tal sistema consistiria de um método de otimização mais geral que poderia otimizar muitos parâmetros de redes ópticas de uma só vez, encontrando as melhores topologias para as redes ópticas, de acordo com os parâmetros considerados. Como é possível que o atendimento aos requerimentos de interesse gerem certo conflito, essa linha de trabalho poderia ajudar a decidir e a balancear quais características priorizar mais (ou menos) no projeto de topologia de redes ópticas.

Há ainda o interesse em inspecionar mais a fundo se as redes de saída das heurísticas fornecidas no presente trabalho, que minimizaram o requerimento de comprimentos de onda, também geram de fato vantagem quanto à sua confiabilidade, como é inicialmente constatado nas análises de redes reais. E também se o uso de uma topologia com reduzido requerimento de comprimentos de onda pode ainda melhorar outras propriedades de interesse como a probabilidade de bloqueio, a resiliência, o CAPEX, dentre outros.

Um tópico de interesse a ser trabalhado é considerar a resiliência das redes à falha de diferentes nós ou enlaces, e verificar como isso afeta os parâmetros de rede que se deseja preservar. Conforme citado na Introdução, há também o interesse em finalizar a análise da influência da presença de comunidades no requerimento de comprimentos de onda. Outro ponto de interesse é verificar possibilidades mais acuradas de estimação do requerimento de comprimentos de onda, por exemplo, por meio de regressões não-lineares. Há a intenção ainda de comparar as heurísticas apresentadas neste trabalho com outras heurísticas globais de otimização para projeto de redes ópticas. Um fator de interesse ainda, é trabalhar com

a *feature extraction* no lugar da *feature selection*, para criar novos invariantes a partir dos existentes, e avaliar eventuais ganhos na explicação do requerimento de comprimentos de onda, e de outros invariantes a serem analisados no futuro.

Um ponto a ser explorado é verificar a importância das invariantes no caso da análise de uma amostra de redes (grafos) com efeito *small-world* ou *scale-free* (WANG; CHEN, 2003), ou ainda redes com uma composição destas duas características, pois muitas vezes as redes reais também apresentam esses perfis (ARAÚJO, 2015), ao invés do comportamento aleatório de redes geradas pelo modelo Erdos-Renyi, usado no presente trabalho.

O estimador de informação mútua utilizado no presente trabalho, baseado no método da coincidência, deixa muito espaço para aplicação em quaisquer outros campos que demandem estimação de entropia; e também trabalhos futuros poderiam compará-lo com estimadores similares da literatura.

Por fim, mas não menos importante, quanto a busca por topologias de redes, nas heurísticas planeja-se adicionar a questão da distância física dos enlaces, e ainda o impedimento da existência de certos enlaces por questões geográficas, para a obtenção de topologias físicas de redes ópticas otimizada, via heurística, além de também considerar os custos de infra-estrutura. Há um igual interesse em considerar a confiabilidade da rede, para que a rede gerada possa ser tão resistente a falhas quanto possível. Outro ponto de interesse ainda é executar simulações para verificar o desempenho das topologias de saída otimizadas como redes do tipo EON, com diferentes esquemas de particionamento espectral.

Acredita-se que essas propostas de trabalhos futuros possam agregar qualidade à rede óptica final obtida, já comparativamente ao que é apresentado no presente trabalho.

Referências

- ADAMIC, L. A.; ADAR, E. Friends and neighbors on the web. *Social networks*, Elsevier, v. 25, n. 3, p. 211–230, 2003. Citado na página 109.
- ANTUNES, C. H.; CRAVEIRINHA, J.; CLIMACO, J. N. A multiple criteria model for new telecommunication service planning. *European journal of operational research*, Elsevier, v. 71, n. 3, p. 341–352, 1993. Citado na página 16.
- ARACIL, J. *et al.* Research in optical burst switching within the e-photon/one network of excellence. *Optical Switching and Networking*, Elsevier, v. 4, n. 1, p. 1–19, 2007. Citado na página 15.
- ARAÚJO, D. R. *et al.* An efficient multi-objective evolutionary optimizer to design all-optical networks considering physical impairments and capex. In: IEEE. *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*. [S.l.], 2011. p. 76–81. Citado 3 vezes nas páginas 24, 25 e 26.
- ARAÚJO, D. R. B. d. *Planejamento de redes ópticas usando inteligência computacional e ciência das redes*. Tese (Doutorado) — Universidade Federal de Pernambuco, 2015. Citado na página 92.
- ARAÚJO, D. R. D.; BASTOS-FILHO, C. J.; MARTINS-FILHO, J. F. An evolutionary approach with surrogate models and network science concepts to design optical networks. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 43, p. 67–80, 2015. Citado 3 vezes nas páginas 17, 24 e 25.
- BANERJEE, D.; MUKHERJEE, B. Wavelength-routed optical networks: linear formulation, resource budgeting tradeoffs, and a reconfiguration study. *IEEE/ACM Trans. Netw.*, IEEE Press, v. 8, n. 5, p. 598–607, 2000. Citado na página 13.
- BANERJEE, N.; KUMAR, R. Multiobjective network design for realistic traffic models. In: ACM. *Proceedings of the 9th annual conference on Genetic and evolutionary computation*. [S.l.], 2007. p. 1904–1911. Citado na página 24.
- BARONI, S.; BAYVEL, P. Wavelength requirements in arbitrarily connected wavelength-routed optical networks. *Journal of Lightwave Technology*, IEEE, v. 15, n. 2, p. 242–251, 1997. Citado 4 vezes nas páginas 21, 33, 53 e 54.
- BARRAT, A. *et al.* The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, National Acad Sciences, v. 101, n. 11, p. 3747–3752, 2004. Citado 2 vezes nas páginas 104 e 106.
- BASTOS-FILHO, C. J. *et al.* Design of transparent optical networks considering physical impairments, capex and energy consumption. In: IEEE. *Transparent Optical Networks (ICTON), 2011 13th International Conference on*. [S.l.], 2011. p. 1–4. Citado na página 25.
- BENNASAR, M.; HICKS, Y.; SETCHI, R. Feature selection using joint mutual information maximisation. *Expert Systems with Applications*, Elsevier, v. 42, n. 22, p. 8520–8532, 2015. Citado 5 vezes nas páginas 17, 18, 32, 58 e 89.

- BERMINGHAM, M. L. *et al.* Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Scientific reports*, Nature Publishing Group, v. 5, p. 10312, 2015. Citado na página 31.
- BLUM, A. L.; LANGLEY, P. Selection of relevant features and examples in machine learning. *Artificial intelligence*, Elsevier, v. 97, n. 1-2, p. 245–271, 1997. Citado na página 30.
- BONACICH, P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, Taylor & Francis, v. 2, n. 1, p. 113–120, 1972. Citado na página 108.
- BONACICH, P. Power and centrality: a family of measures. *American journal of sociology*, University of Chicago Press, v. 92, n. 5, p. 1170–1182, 1987. Citado na página 108.
- BONACICH, P.; LLOYD, P. Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, Elsevier, v. 23, n. 3, p. 191–201, 2001. Citado na página 108.
- BRANDES, U. *et al.* On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 20, n. 2, p. 172–188, 2008. Citado na página 163.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, Elsevier, v. 30, n. 1, p. 107–117, 1998. Citado 2 vezes nas páginas 108 e 109.
- BURKE, E. K.; KENDALL, G. *et al.* *Search methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. [S.l.]: Springer, 2014. Citado na página 68.
- BURT, R. S. Structural holes and good ideas. *American journal of sociology*, JSTOR, v. 110, n. 2, p. 349–399, 2004. Citado na página 108.
- CAPOROSSI, G. *et al.* Centrality and betweenness: vertex and edge decomposition of the Wiener index. *Match-Communications in Mathematical and Computer Chemistry*, v. 68, n. 1, 2012. Citado 5 vezes nas páginas 19, 20, 105, 106 e 107.
- CHÂTELAIN, B. *et al.* Topological wavelength usage estimation in transparent wide area networks. *Journal of Optical Communications and Networking*, Optical Society of America, v. 1, n. 1, p. 196–203, 2009. Citado 4 vezes nas páginas 22, 49, 55 e 105.
- CHAVES, D. A.; BASTOS-FILHO, C. J.; MARTINS-FILHO, J. F. Multiobjective physical topology design of all-optical networks considering QoS and Capex. In: IEEE. *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*. [S.l.], 2010. p. 1–3. Citado 3 vezes nas páginas 24, 25 e 26.
- CHRISTODOULOPOULOS, K.; TOMKOS, I.; VARVARIGOS, E. Elastic bandwidth allocation in flexible OFDM-based optical networks. *Journal of Lightwave Technology*, IEEE, v. 29, n. 9, p. 1354–1366, 2011. Citado 2 vezes nas páginas 16 e 24.
- CHUNG, F. R. *Spectral graph theory*. [S.l.]: American Mathematical Soc., 1997. Citado 2 vezes nas páginas 20 e 64.

- CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. *Physical review E*, APS, v. 70, n. 6, p. 066111/1–066111/6, 2004. Citado 4 vezes nas páginas 20, 107, 160 e 163.
- COHEN, J. *Statistical power analysis for the behavioral sciences*. [S.l.]: Erlbaum Associates, Hillsdale, 1987. Citado 3 vezes nas páginas 63, 151 e 152.
- COUSINEAU, M. *et al.* RWA problem with geodesics in realistic OTN topologies. *Optical Switching and Networking*, Elsevier, v. 15, p. 18–28, 2015. Citado 11 vezes nas páginas 13, 15, 23, 34, 36, 58, 66, 77, 87, 90 e 175.
- COVER, T. M.; THOMAS, J. A. *Elements of information theory*. [S.l.]: John Wiley & Sons, 2012. Citado na página 39.
- CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. *InterJournal, Complex Systems*, v. 1695, n. 5, p. 1–9, 2006. Disponível em: <<http://igraph.org>>. Citado 10 vezes nas páginas 20, 35, 72, 104, 105, 106, 107, 108, 109 e 164.
- CVETKOVIC, D. M.; DOOB, M.; SACHS, H. *Spectra of Graphs: Theory and Applications*. [S.l.]: New York: Wiley, 1998. Citado na página 104.
- DEPIZZOL, D. B. *et al.* Feature selection for optical network design via a new mutual information estimator. *Expert Systems with Applications*, Elsevier, v. 107, p. 72–88, 2018. Citado 2 vezes nas páginas 17 e 36.
- DHARMAWEERA, N.; PARTHIBAN, R.; SEKERCIOGLU, Y. Multi-constraint physical topology design for all optical networks. In: IEEE. *Telecommunications (ICT), 2011 18th International Conference on*. [S.l.], 2011. p. 463–469. Citado 2 vezes nas páginas 24 e 25.
- DIESTEL, R. *Graph theory*. [S.l.]: Springer, 2016. v. 173. Citado 5 vezes nas páginas 18, 19, 20, 103 e 104.
- DUPLEIX, V.; ARAÚJO, D. R.; BASTOS-FILHO, C. J. Designing the optical network of Haiti using a multi-objective evolutionary approach. In: IEEE. *Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on*. [S.l.], 2016. p. 1–6. Citado na página 24.
- EAGLE, N.; MACY, M.; CLAXTON, R. Network diversity and economic development. *Science*, American Association for the Advancement of Science, v. 328, n. 5981, p. 1029–1031, 2010. Citado na página 108.
- EPPSTEIN, D.; LÖFFLER, M.; STRASH, D. Listing all maximal cliques in sparse graphs in near-optimal time. In: SPRINGER. *International Symposium on Algorithms and Computation*. Jeju Island, Korea, 2010. p. 403–414. Citado na página 105.
- ERDÖS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, Citeseer, v. 5, p. 17–61, 1960. Citado na página 33.
- ESTRADA, E. The communicability distance in graphs. *Linear Algebra and its Applications*, Elsevier, v. 436, n. 11, p. 4317–4328, 2012. Citado 2 vezes nas páginas 19 e 106.

- ESTRADA, E.; RODRIGUEZ-VELAZQUEZ, J. A. Subgraph centrality in complex networks. *Physical Review E*, APS, v. 71, n. 5, p. 056103/1–056103/9, 2005. Citado na página 109.
- FENGER, C. *et al.* Statistical study of the correlation between topology and wavelength usage in optical networks with and without conversion. In: *Networking 2000 Broadband Communications, High Performance Networking, and Performance of Communication Networks*. [S.l.]: Springer, 2000. p. 168–175. Citado 4 vezes nas páginas 21, 49, 55 e 64.
- FIEDLER, M. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, v. 23, n. 98, p. 298–305, 1973. Citado na página 104.
- FOX, J. *Applied regression analysis and generalized linear models*. [S.l.]: Sage Publications, 2015. Citado 2 vezes nas páginas 63 e 151.
- FREEMAN, L. C. Centrality in social networks conceptual clarification. *Social networks*, Elsevier, v. 1, n. 3, p. 215–239, 1979. Citado 5 vezes nas páginas 20, 21, 104, 106 e 107.
- FRISKNEY, R.; DONG, S.; PHILLIPS, C. *Differentiated resilience in optical networks*. [S.l.]: Google Patents, 2004. US Patent App. 10/323,409. Citado na página 17.
- GROSS, J. L.; YELLEN, J.; ZHANG, P. *Handbook of graph theory*. [S.l.]: Chapman and Hall/CRC, 2013. Citado na página 65.
- GROVER, W. D.; DOUCETTE, J. Topological design of survivable mesh-based transport networks. *Annals of operations research*, Springer, v. 106, n. 1-4, p. 79–125, 2001. Citado na página 24.
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of machine learning research*, v. 3, n. Mar, p. 1157–1182, 2003. Citado 2 vezes nas páginas 30 e 31.
- GUYON, I.; ELISSEEFF, A. An introduction to feature extraction. In: *Feature extraction*. [S.l.]: Springer, 2006. p. 1–25. Citado 2 vezes nas páginas 31 e 58.
- GUYON, I. *et al.* *Feature extraction: foundations and applications*. [S.l.]: Springer, 2008. v. 207. Citado na página 30.
- HAGBERG, A.; SWART, P.; CHULT, D. S. *Exploring network structure, dynamics, and function using NetworkX*. [S.l.], 2008. Citado na página 33.
- HAIGHT, F. A. *Handbook of the poisson distribution*. Wiley, 1967. Citado na página 63.
- HAN, Z.; ZHANG, K. *Surrogate-based optimization, Real-World Applications of Genetic Algorithms*. [S.l.], 2012. Citado na página 25.
- HANAY, Y. S.; ARAKAWA, S.; MURATA, M. Network topology selection with multistate neural memories. *Expert Systems with Applications*, Elsevier, v. 42, n. 6, p. 3219–3226, 2015. Citado 2 vezes nas páginas 16 e 89.
- HANSEN, P.; MLADENOVIC, N. Variable neighborhood search. In: *Search methodologies*. [S.l.]: Springer, 2014. p. 313–337. Citado na página 74.

- HU, Q. *et al.* Measuring relevance between discrete and continuous features based on neighborhood mutual information. *Expert Systems with Applications*, Elsevier, v. 38, n. 9, p. 10737–10750, 2011. Citado 3 vezes nas páginas 17, 18 e 89.
- IZQUIERDO-ZARAGOZA, J.-L.; PAVON-MARINO, P.; BUENO-DELGADO, M.-V. Distance-adaptive online RSA algorithms for heterogeneous flex-grid networks. In: IEEE. *Optical Network Design and Modeling, 2014 International Conference on*. [S.l.], 2014. p. 204–209. Citado 2 vezes nas páginas 14 e 174.
- JAMES, G. *et al.* *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112. Citado na página 30.
- JAUMARD, B.; MEYER, C.; THIONGANE, B. Comparison of ILP formulations for the RWA problem. *Optical Switching and Networking*, Elsevier, v. 4, n. 3, p. 157–172, 2007. Citado na página 13.
- JOYCE, K. E. *et al.* A new measure of centrality for brain networks. *PLoS One*, Public Library of Science, v. 5, n. 8, p. e12200/1–e12200/13, 2010. Citado na página 107.
- JUN, W. *et al.* Natural connectivity of complex networks. *Chinese physics letters*, IOP Publishing, v. 27, n. 7, p. 078902–1 – 078902–1, 2010. Citado na página 104.
- KANEVSKY, A. Finding all minimum-size separating vertex sets in a graph. *Networks*, Wiley Online Library, v. 23, n. 6, p. 533–541, 1993. Citado na página 105.
- KEL’MANS, A. K. Connectivity of probabilistic networks. *Automation and Remote Control*, v. 3, p. 98–116, 1967. Citado na página 84.
- KERNIGHAN, B. W.; RITCHIE, D. M. *The C programming language*. [S.l.: s.n.], 2006. Citado na página 33.
- KHURI, A. I.; MUKHOPADHYAY, S. Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*, Wiley Online Library, v. 2, n. 2, p. 128–149, 2010. Citado na página 61.
- KLEINBERG, J. M. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, ACM, v. 46, n. 5, p. 604–632, 1999. Citado na página 108.
- KOENKER, R. *Quantile Regression*. [S.l.]: Cambridge University Press, New York, 2005. Citado na página 63.
- KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. *Artificial intelligence*, Elsevier, v. 97, n. 1-2, p. 273–324, 1997. Citado na página 30.
- LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. *Physical review letters*, APS, v. 87, n. 19, p. 198701/1–198701/4, 2001. Citado na página 106.
- LIU, H.; TOBAGI, F. A. Physical topology design for all-optical networks. *Optical Switching and Networking*, Elsevier, v. 5, n. 4, p. 219–231, 2008. Citado na página 24.
- LIU, W.; PAWLIKOWSKI, K.; SIRISENA, H. Algebraic connectivity metric for spare capacity allocation problem in survivable networks. *Computer Communications*, Elsevier, v. 34, n. 12, p. 1425–1435, 2011. Citado na página 104.

- LU, L.; ZHANG, M. Edge betweenness centrality. In: *Encyclopedia of systems biology*. [S.l.]: Springer, 2013. p. 647–648. Citado na página 64.
- LUKOVITS, I.; NIKOLIĆ, S.; TRINAJSTIĆ, N. Resistance distance in regular graphs. *International journal of quantum chemistry*, Wiley Online Library, v. 71, n. 3, p. 217–225, 1999. Citado 2 vezes nas páginas 20 e 105.
- MA, S. Calculation of entropy from data of motion. *Journal of Statistical Physics*, Springer, v. 26, n. 2, p. 221–240, 1981. Citado 2 vezes nas páginas 32 e 37.
- MA, S. *Statistical Mechanics*. [S.l.]: World Scientific Publishing Co. Pte. Ltd., 1985. Citado 3 vezes nas páginas 18, 32 e 37.
- MARSAN, M. A. *et al.* Topologies for wavelength-routing all-optical networks. *IEEE/ACM Transactions on Networking (TON)*, IEEE Press, v. 1, n. 5, p. 534–546, 1993. Citado na página 15.
- MCKAY, B. D. *Nauty User's Guide (Version 2.2)*. [S.l.]: Technical Report TR-CS-9002, Australian National University, 2003. Citado na página 23.
- MONTALVÃO, J.; ATTUX, R.; SILVA, D. A pragmatic entropy and differential entropy estimator for small datasets. *Journal of Communication and Information Systems*, Public Knowledge Project, v. 29, n. 1, p. 29–36, 2014. Citado 4 vezes nas páginas 17, 32, 40 e 89.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. [S.l.]: John Wiley & Sons, 2012. v. 821. Citado na página 30.
- MORAIS, R. M. *et al.* Genetic algorithm for the topological design of survivable optical transport networks. *Journal of optical communications and networking*, Optical Society of America, v. 3, n. 1, p. 17–26, 2011. Citado na página 24.
- NEWMAN, M. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, APS, v. 74, n. 3, p. 036104(1)–036104(22), 2006. Citado na página 163.
- NEWMAN, M.; GIRVAN, M. Finding and evaluating community structure in networks. *Physical Review E*, APS, v. 69, n. 2, p. 026113(1)–026113(16), 2004. Citado 2 vezes nas páginas 160 e 163.
- NEWMAN, M. E. Assortative mixing in networks. *Physical review letters*, APS, v. 89, n. 20, p. 208701/1–208701/4, 2002. Citado na página 104.
- NEWMAN, M. E. Mixing patterns in networks. *Physical Review E*, APS, v. 67, n. 2, p. 026126/1–026126/13, 2003. Citado na página 104.
- PALACIOS, J. L. Closed-form formulas for kirchhoff index. *International Journal of Quantum Chemistry*, Wiley Online Library, v. 81, n. 2, p. 135–140, 2001. Citado 2 vezes nas páginas 20 e 105.
- PATEL, A. N. *et al.* Routing, wavelength assignment, and spectrum allocation algorithms in transparent flexible optical WDM networks. *Optical Switching and Networking*, Elsevier, v. 9, n. 3, p. 191–204, 2012. Citado na página 16.

- PAVAN, C. *et al.* How reliable are the real-world optical transport networks? *Journal of Optical Communications and Networking*, Optical Society of America, v. 7, n. 6, p. 578–585, 2015. Citado 3 vezes nas páginas 17, 84 e 85.
- PAVAN, C. *et al.* Generating realistic optical transport network topologies. *Journal of Optical Communications and Networking*, Optical Society of America, v. 2, n. 1, p. 80–90, 2010. Citado na página 33.
- PLAVŠIĆ, D. *et al.* On the Harary index for the characterization of chemical graphs. *Journal of Mathematical Chemistry*, Springer, v. 12, n. 1, p. 235–250, 1993. Citado 2 vezes nas páginas 20 e 105.
- PONS, P.; LATAPY, M. Computing communities in large networks using random walks. In: SPRINGER. *International Symposium on Computer and Information Sciences*. [S.l.], 2005. p. 284–293. Citado na página 163.
- PROVAN, J. S.; BALL, M. O. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal on Computing*, SIAM, v. 12, n. 4, p. 777–788, 1983. Citado na página 85.
- PRZEWOŹNICZEK, M. *et al.* Towards solving practical problems of large solution space using a novel pattern searching hybrid evolutionary algorithm—an elastic optical network optimization case study. *Expert Systems with Applications*, Elsevier, v. 42, n. 21, p. 7781–7796, 2015. Citado 2 vezes nas páginas 16 e 89.
- QUEIPO, N. V. *et al.* Surrogate-based analysis and optimization. *Progress in aerospace sciences*, Elsevier, v. 41, n. 1, p. 1–28, 2005. Citado 2 vezes nas páginas 25 e 61.
- R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria, 2016. Disponível em: <<http://www.R-project.org/>>. Citado 2 vezes nas páginas 35 e 164.
- RAGHAVAN, U.; ALBERT, R.; KUMARA, S. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, APS, v. 76, n. 3, p. 036106(1)–036106(12), 2007. Citado na página 163.
- ROSVALL, M.; AXELSSON, D.; BERGSTROM, C. T. The map equation. *The European Physical Journal-Special Topics*, Springer, v. 178, n. 1, p. 13–23, 2009. Citado na página 163.
- SHEELA, D. *Dimensioning strategies using topological parameters in optical mesh networks*. Tese (Doutorado) — FACULTY OF INFORMATION AND COMMUNICATION ENGINEERING, ANNA UNIVERSITY, Chennai, Índia, 2012. Citado 2 vezes nas páginas 15 e 21.
- SHI, J.; FONSEKA, J. P. Hierarchical self-healing rings. *IEEE/ACM Transactions on Networking (TON)*, IEEE Press, v. 3, n. 6, p. 690–697, 1995. Citado na página 161.
- SIMMONS, J. M. *Optical network design and planning*. [S.l.]: Springer, 2014. Citado 2 vezes nas páginas 24 e 161.
- SKIENA, S. *Implementing discrete mathematics: Combinatorics and graph theory with mathematica*. [S.l.]: Basic Books, 1990. Citado 2 vezes nas páginas 19 e 106.

TALEBI, S. *et al.* Spectrum management techniques for elastic optical networks: a survey. *Optical Switching and Networking*, Elsevier, v. 13, p. 34–48, 2014. Citado na página 16.

TESSINARI, R. E. *A Fairness-Focused Spectrum Assignment Algorithm For Elastic Optical Networks*. Tese (Doutorado) — Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil, 2016. Citado 2 vezes nas páginas 14 e 174.

TESSINARI, R. S. *et al.* Zone based spectrum assignment in elastic optical networks: a fairness approach. In: IEEE. *Opto-Electronics and Communications Conference (OECC), 2015*. [S.l.], 2015. p. 1–3. Citado na página 25.

TESSINARI, R. S. *et al.* ElasticO++: an elastic optical network simulation framework for OMNeT++. *Optical Switching and Networking*, Elsevier, v. 22, p. 95–104, 2016. Citado na página 16.

TRAAG, V.; BRUGGEMAN, J. Community detection in networks with positive and negative links. *Physical Review E*, APS, v. 80, n. 3, p. 036115(1)–036115(7), 2009. Citado na página 163.

VELASCO, L. *et al.* Modeling the routing and spectrum allocation problem for flexgrid optical networks. *Photonic Network Communications*, Springer, v. 24, n. 3, p. 177–186, 2012. Citado na página 16.

WANG, R.; MUKHERJEE, B. Spectrum management in heterogeneous bandwidth optical networks. *Optical Switching and Networking*, Elsevier, v. 11, p. 83–91, 2014. Citado 2 vezes nas páginas 16 e 25.

WANG, X. F.; CHEN, G. Complex networks: small-world, scale-free and beyond. *IEEE circuits and systems magazine*, IEEE, v. 3, n. 1, p. 6–20, 2003. Citado 2 vezes nas páginas 22 e 92.

WASSERMAN, S.; FAUST, K. *Social network analysis: methods and applications*. [S.l.]: Cambridge university press, 1994. v. 8. Citado 4 vezes nas páginas 21, 104, 106 e 107.

WASSERSTEIN, R. L.; LAZAR, N. A. *The ASA’s statement on p-values: context, process, and purpose*. [S.l.]: Journal The American Statistician, Taylor & Francis, 2016. 129–133 p. Citado na página 63.

WHITE, D. R.; HARARY, F. The cohesiveness of blocks in social networks: Node connectivity and conditional density. *Sociological Methodology*, JSTOR, p. 305–359, 2001. Citado na página 105.

WU, J. *et al.* Robustness of regular graphs based on natural connectivity. *ArXiv e-prints*, 2009. Citado na página 104.

XIAO, G.; LEUNG, Y.-W.; HUNG, K.-W. Two-stage cut saturation algorithm for designing all-optical networks. *IEEE Transactions on Communications*, IEEE, v. 49, n. 6, p. 1102–1115, 2001. Citado 2 vezes nas páginas 24 e 25.

YANG, X.; RAMAMURTHY, B. Dynamic routing in translucent wdm optical networks: the intradomain case. *Journal of Lightwave Technology*, v. 23, n. 3, p. 955–971, 2005. Citado na página 13.

- YANG, Y. *et al.* Multi-objective optimization based on ant colony optimization in grid over optical burst switching networks. *Expert Systems with Applications*, Elsevier, v. 37, n. 2, p. 1769–1775, 2010. Citado 2 vezes nas páginas 16 e 89.
- YIN, Y. *et al.* Fragmentation-aware routing, modulation and spectrum assignment algorithms in elastic optical networks. In: IEEE. *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*. [S.l.], 2013. p. 1–3. Citado na página 25.
- YOON, M.; BAEK, Y.; TCHA, D. Design of a distributed fiber transport network with hubbing topology. *European Journal of Operational Research*, Elsevier, v. 104, n. 3, p. 510–520, 1998. Citado na página 16.
- YUAN, P.; XU, A. The influence of physical network topologies on wavelength requirements in optical networks. *Journal of Lightwave Technology*, IEEE, v. 28, n. 9, p. 1338–1343, 2010. Citado 2 vezes nas páginas 22 e 55.
- ZANG, H.; JUE, J. P.; MUKHERJEE, B. A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. *Optical Networks Magazine*, v. 1, n. 1, p. 47–60, 2000. Citado 2 vezes nas páginas 13 e 14.
- ZHANG, G. *et al.* A survey on OFDM-based elastic core optical networking. *IEEE Communications Surveys & Tutorials*, IEEE, v. 15, n. 1, p. 65–87, 2013. Citado na página 16.

Apêndices

APÊNDICE A – Notações e Definições

LEGENDA:

M = Medida

I = Invariante

Name	Sym- bol	Description	Computation Method
Graph	$G(V, E)$ or G	A graph is denoted by $G(V, E)$, or simply G , where V is a set of vertices and E is a set of edges (DIESTEL, 2016), where edges connect vertices.	(not applicable)
Vertices number (I)	n	It is the number of vertices in a graph $G(V, E)$ (DIESTEL, 2016), also called order of the graph.	$n = V $
Edges number (I)	m	It is the number of edges in a graph $G(V, E)$ (DIESTEL, 2016), also called size of the graph.	$m = E $
Path	P	In a graph, a path P interconnecting a pair of vertices $\{i, j\}$ is a sequence of edges $\{(i, v_1), (v_1, v_2), \dots, (v_{\psi-1}, v_{\psi}), (v_{\psi}, j)\}$ which connect a sequence of vertices $\{i, v_1, v_2, \dots, v_{\psi}, j\}$ (DIESTEL, 2016).	(not applicable)
Minimum path or Geodesic	P_{min}	In a graph, a minimum path interconnecting a pair of vertices $\{i, j\}$ is a path P with the minimum size, i.e., with the minimum number of edges (DIESTEL, 2016). In the literature, minimum path is also called geodesic.	(not applicable)
Distance (M)	d	The distance between two vertices i and j is given by the number of edges in a geodesic P_{min} interconnecting the pair $\{i, j\}$ (DIESTEL, 2016). All values of $P_{min}(i, j)$, from all pairs $i, j \in V$, compose the distance matrix $d(i, j)$.	(description is enough for computation)
Diameter (I)	D	It is the greatest value of distance $d(i, j)$ for all pair of vertices from a graph G (DIESTEL, 2016).	$D = \max\{d(i, j)\}; i, j \in V.$
Adjacency matrix	$A(i, j)$	A is a matrix representation of a graph and used to compute many graph properties (DIESTEL, 2016).	$A(i, j)$ assumes the value 1 in the position (i, j) if there is an edge (i, j) in the graph G , and the value 0 otherwise (DIESTEL, 2016).
Vertex degree (M)	deg	The vertex degree, or simply degree, of a vertex i is the number of edges connected to i (DIESTEL, 2016).	$deg(i) = \sum_{j=1}^n A(i, j)$; where A is the adjacency matrix.
Normalized vertex degree (M)	deg_n	The normalized degree of a vertex i is the number of edges connected to i (DIESTEL, 2016) divided by $n - 1$.	$deg_n(i) = deg(i)/(n-1)$; where $i \in V$.

Continues on the next page...

...continued from previous page.			
Ratio vertices with minimum degree (I)	R_{min}	It is the proportion of vertices in the graph that has the minimum degree.	$R_{min} = V[deg_{min}] /n$, where $V[deg_{min}]$ is the set of vertices that have the minimum degree δ .
Ratio vertices with maximum degree (I)	R_{max}	It is the proportion of vertices in the graph that has the maximum degree.	$R_{max} = V[deg_{max}] /n$, where $V[deg_{max}]$ is the set of vertices that have the maximum degree Δ .
Degree centralization (I)	deg_C	It is a measure of centrality in a graph level, wich uses the degrees of vertices.	$deg_C = \frac{\sum_{i=1}^n [\Delta - deg(i)]}{deg_{C(max)}}$; where $deg_{C(max)}$ is a maximal theoretical of deg_C (Freeman (1979), Wasserman e Faust (1994) apud Csardi e Nepusz (2006)).
Normalized degree centralization (I)	deg_{C_n}	It is a measure of centrality in a graph level, wich uses the normalized degree of vertices.	$deg_{C_n} = \frac{\sum_{i=1}^n [\Delta - deg_n(i)]}{deg_{C_n(max)}}$; where $deg_{C_n(max)}$ is a maximal theoretical of deg_{C_n} (Freeman (1979), Wasserman e Faust (1994) apud Csardi e Nepusz (2006)).
Average nearest neighbor degree (M)	ND_i	Compute the average nearest neighbor degree of each vertex i .	$ND_i = \left(\sum_{j=1}^k deg(j) \right) / k$, where the vertices $1, \dots, k$ are neighbors of the vertex i (Barrat et al. (2004) apud Csardi e Nepusz (2006)).
Assortativity degree (I)	r	The assortativity degree is positive if vertices with high degrees tend to connect with each other, and negative otherwise.	$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k)$, where e_{jk} is the fraction of edges that connects vertices with degree j and k , $q_i = \sum_j e_{ij}$, and σ_q^2 is the variance of q_i (Newman (2002), Newman (2003) apud Csardi e Nepusz (2006)).
Diagonal matrix	Diag	Diag is a matrix which assumes the values of the degrees in its diagonal.	$Diag(i, i) = deg(i)$, to $i \in \{1, \dots, n\}$ and $Diag(i, j) = 0$, for all pair $\{i, j\}$ with $i \neq j$.
Laplacian matrix	L	Laplacian matrix is a matrix representation of a graph, and can be used to find many graph properties (CVETKOVIC; DOOB; SACHS, 1998).	$L(i, j) = Diag(i, j) - A(i, j)$ (CVETKOVIC; DOOB; SACHS, 1998).
Algebraic connectivity (I)	μ_2	According to Liu, Pawlikowski e Sirlsena (2011), this measure represents the network potential immunity to failure and thus it is important in analysis of network survivability.	μ_2 is the second smallest eigenvalue of laplacian matrix L (FIEDLER, 1973).
Natural connectivity (I)	$\bar{\beta}$	It provides a quantitative analysis of network robustness (WU et al., 2009).	$\bar{\beta} = \ln(\sum_{i=1}^n e^{\beta_i} / n)$; where $\beta_1, \beta_2, \dots, \beta_n$ are the eigenvalues of the adjacency matrix A (JUN et al., 2010).
Edge density (I)	α	It is the ratio between the edges number (m) and the largest possible number of edges $n(n-1)/2$ in a graph $G(V, E)$, with $ V =n$ and $ E =m$ (DIESTEL, 2016).	$\alpha = 2m/(n(n - 1))$

Continues on the next page...

...continued from previous page.			
Vertex connectivity (I)	K	The vertex connectivity of a pair of vertices in a graph G is the minimum number of vertices needed to break all disjoint paths which connect them, or to reduce the G to a vertex. The vertex connectivity of a graph is the minimum of the vertex connectivity of every pair of vertices in the graph (White e Harary (2001) apud Csardi e Nepusz (2006)).	(description is enough for computation)
Edge connectivity (I)	K'	The edge connectivity of a pair of vertices can be interpreted as the minimum number of edges needed to break all disjoint paths which connect them. The edge connectivity of a graph is the minimum of the edge connectivity of every pair of vertices in the graph (White e Harary (2001) apud Csardi e Nepusz (2006)).	(description is enough for computation)
Number of minimum sets that disconnects the graph (I)	NMSD	The number of vertex sets of minimum size whose removal separates the graph into two or more components.	All these sets are found via Kanevsky algorithm (Kanevsky (1993) apud Csardi e Nepusz (2006)), and then it is simply counted how many sets are found.
Clique	C	A clique is a complete subgraph in a graph.	Cliques are obtained using the algorithm of Eppstein, Löffler e Strash (2010).
Proportion of largest clique (I)	PLC	It shows how close is the graph to be a complete graph.	PLC = $ C_L /n$, where C_L is a largest clique obtained using the algorithm of Eppstein, Löffler e Strash (2010).
Amount of largest cliques (I)	ALC	It shows how many different ways there are cliques with the order of the largest clique.	It is the number of largest cliques in the graph, where the cliques are obtained using the algorithm of Eppstein, Löffler e Strash (2010).
Number of spanning trees (I)	St	A tree is a connected undirected graph with no cycles. A spanning tree of a graph $G(V, E)$ is a subgraph of G which connects all vertices V with $n - 1$ vertices from E (CHÂTELAIN <i>et al.</i> , 2009).	By Châtelain <i>et al.</i> (2009), $St = \det(L[-i, -j]) $, where $L[-i, -j]$ is the laplacian matrix without any line i and any column j .
Kirchhoff index (I)	K_f	The Kirchhoff index (LUKOVITS; NIKOLIĆ; TRINAJSTIĆ, 1999; PALACIOS, 2001) is half the sum of the resistance distance matrix Ω_{ij} .	$K_f = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Omega_{ij},$ where for non-directed graphs $\Omega_{ij} = (\Gamma)_{ii}^{-1} + (\Gamma)_{jj}^{-1} - 2(\Gamma)_{ij}^{-1}$, with Γ given by $\Gamma = L + [1/n]_n$, where L is the Laplacian matrix, and $[1/n]_n$ is a $n \times n$ matrix consisting of $1/n$ in all entries.
Wiener index (I)	W	It is half the sum of all distances in a graph (CAPOROSSO <i>et al.</i> , 2012).	$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d(i, j); i, j \in V.$
Harary index (I)	H	The Harary index (PLAVŠIĆ <i>et al.</i> , 1993) of a graph is half the sum of the reciprocal of the graph distance matrix $(1/d(i, j))$.	$H = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{d(i, j)}; i, j \in V.$
Continues on the next page...			

...continued from previous page.			
Transmission (M)	T_i	Transmission T_i of the vertex i is the sum of all distances from i to all other vertices (CAPOROSS <i>et al.</i> , 2012).	$T_i = \sum_{j=1}^n d(i, j); i, j \in V.$
Communicability distance (M)	C_d	C_d is an option of Euclidean distance in graphs (ESTRADA, 2012).	$C_d[p, q] = G_{pp} + G_{qq} - 2G_{pq}$ with $G_{pq} = \sum_{j=1}^n \varphi_j(p)\varphi_j(q)e^{\lambda_j}$, where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ are the eigenvalues of the adjacency matrix A, and where $\varphi_j(p)$ and $\varphi_j(q)$ are the p^{th} and q^{th} entries of the j^{th} orthonormal eigenvector of A associated with the eigenvalue λ_j .
Adjusted Communicability distance (M)	AC_d	AC_d is an measure inspired on Communicability distance (C_d) (ESTRADA, 2012).	$AC_d[p, q] = G_{pp} + G_{qq} - 2G_{pq}$ with $G_{pq} = exp(A_{pq})$, where A is the adjacency matrix.
Communication efficiency (I)	C_e	We can say that the distance is inversely proportional to the efficiency. Based on this, we can define that the average of the inverse of the distances is the C_e (LATORA; MARCHIORI, 2001).	$C_e = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d(i, j)} ; i, j \in V.$
Adjacent triangles (M)	AT_i	Count how many triangles a vertex i is part of, in a graph (CSARDI; NEPUSZ, 2006).	(description is enough for computation)
Global transitivity (I)	GT	Global transitivity (or also called global clustering coefficient) measures the probability that the adjacent vertices of a vertex are connected (WASSERMAN e FAUST (1994), BARRAT <i>et al.</i> (2004) <i>apud</i> Csardi e Nepusz (2006)).	A triplet is a set of three vertices that are connected by two or three edges. A triangle is three-closed triplet, one centred on each of the vertices. GT is the simply ratio of the triangles and the connected triplets.
Local transitivity (M)	LT_i	Local transitivity (or also called local clustering coefficient) of a vertex i measures the probability that the adjacent vertices of i are connected (WASSERMAN e FAUST (1994), BARRAT <i>et al.</i> (2004) <i>apud</i> Csardi e Nepusz (2006)).	The local transitivity LT_i of a vertex i is the ratio of triangles connected to the vertex and triplets centered on the vertex.
Eccentricity (M)	E_i	The eccentricity E_i of a vertex i is its maximal shortest path distance from i to all other vertices in the graph (SKLENA, 1990).	$E_i = \max_j [d(i, j)]$
Edge betweenness (M)	EB_{ij}	The edge betweenness EB_{ij} is a measure defined using the number of geodesics going through the edge (i, j) (FREEMAN (1979) <i>apud</i> Csardi e Nepusz (2006)).	$EB_{ij} = \sum g_{kl}^{ij}/g_{kl}; k, l \in V, k < l, \forall (i, j) \in E; g_{kl}^{ij}$ is the number of geodesics that link k and l , passing by (i, j) , and g_{kl} is the number of all geodesics that link k and l .
Vertex betweenness (M)	VB_i	The vertex betweenness VB_i is a measure defined by the number of geodesics going through the vertex i (FREEMAN (1979) <i>apud</i> Csardi e Nepusz (2006)).	$VB_i = \sum g_{kl}^i/g_{kl}; k, l \in V, k < l, \forall i \in V; g_{kl}^i$ is the number of geodesics that link k and l , passing by i , and g_{kl} is the number of all geodesics that link k and l .
Vertex betweenness centrality (I)	VBC	VBC is the vertex betweenness in a graph-level (FREEMAN (1979), WASSERMAN e FAUST (1994) <i>apud</i> Csardi e Nepusz (2006)).	$VBC = \sum_i [\max_j (VB_j) - VB_i]; i, j = 1, \dots, n.$

Continues on the next page...

...continued from previous page.			
Vertex betweenness normalized (M)	VBN_i	VBN_i is a measure that represents how much the vertex i is used, in relation to the most utilized vertex (Freeman (1979) apud Csardi e Nepusz (2006)).	$VBN_i = VB_i/VB_{MaxTeor}$, where $VB_{MaxTeor}$ is the maximum theoretical score for a graph with the same number of vertices and edges.
Vertex betweenness centrality normalized (I)	VBCN	VBCN is the vertex betweenness normalized in a graph-level (Freeman (1979), Wasserman e Faust (1994) apud Csardi e Nepusz (2006)).	$VBCN = \sum_i [\max_j (VBN_j) - VBN_i]$; $i, j = 1, \dots, n$.
Adjusted betweenness centrality (M)	ABC_i	ABC_i is the sum of edge betweenness of all edges linked to a vertex i (CAMPOROSSO <i>et al.</i> , 2012). It refers to the demand of i , such as in VB_i .	$ABC_i = \sum_j EB_{ij}$; $(i, j) \in E$.
Modularity (M)	MB_{ij}	It is defined as the difference of the adjacency matrix and the null model matrix configuration (Clauset, Newman e Moore (2004) apud Csardi e Nepusz (2006)).	$MB_{ij} = A(i, j) - [\deg(i) \deg(j)]/2m$
Global modularity betweenness (I)	GMB	First, a division is made using a community structure detection based on edge betweenness. Then GMB measures how good is this division (Clauset, Newman e Moore (2004) apud Csardi e Nepusz (2006)). Larger GMB values indicate that there are more connections within communities than between them.	$GMB = \frac{1}{2m} \sum_{ij} [A(i, j) - \deg(i) \deg(j)/2m] \delta(i, j)$; with $\delta(i, j) = 1$, if the vertex i and j are in the same community, and $\delta(i, j) = 0$ otherwise. In addition, A is the adjacency matrix, and $\deg(i)$ is the degree of the vertex i .
Leverage centrality (M)	l_i	Leverage centrality is a measure of the relationship between the degree of a given node i and the degree of each of its neighbors N_i (JOYCE <i>et al.</i> , 2010).	$l_i = \frac{1}{\deg(i)} \sum_{j=1}^{ N_i } [\deg(i) - \deg(j)] / [\deg(i) + \deg(j)]$
Closeness centrality (M)	CCV_i	Closeness centrality measures the inverse of how many steps are required to access every other vertex from a given vertex i (Freeman (1979) apud Csardi e Nepusz (2006)).	$CCV_i = \frac{1}{\sum_{j=1}^n d(i, j)}$, with $i \neq j$.
Closeness centrality normalized (M)	$CCVN_i$	$CCVN_i$ is the closeness centrality of vertices normalized, where the normalization is performed by multiplying the closeness of a vertex by $(n - 1)$ (Freeman (1979) apud Csardi e Nepusz (2006)).	$CCVN_i = (n - 1) CCV_i$
Closeness centralization (I)	CC	CC is the closeness centrality of vertices in a graph-level (Freeman (1979), Wasserman e Faust (1994) apud Csardi e Nepusz (2006)).	$CC = \sum_i [\max_j (CCV_j) - CCV_i]$; $i, j = 1, \dots, n$.
Closeness centralization normalized (I)	CCN	CCN is the closeness centrality of vertices normalized in a graph-level (Freeman (1979), Wasserman e Faust (1994) apud Csardi e Nepusz (2006)).	$CCN = \sum_i [\max_j (CCVN_j) - CCVN_i]$; $i, j = 1, \dots, n$.
Continues on the next page...			

...continued from previous page.			
Eigenvector centrality by vertices (M)	ECV_i	The centrality of each vertex is proportional to the sum of the centralities of its vertices neighbors (Bonacich (1987) <i>apud</i> Csardi e Nepusz (2006)).	First, we compute the eigenvalues of the adjacency matrix A of the graph, and after we sort them. The eigenvector, that corresponds to the biggest eigenvalue, is the vector ECV_i .
Eigenvector centrality global (I)	ECG	ECG is the eigenvalue corresponding to the computed eigenvector ECV_i , which contains the centrality scores (Bonacich (1987) <i>apud</i> Csardi e Nepusz (2006)).	We compute the eigenvalues of the adjacency matrix A of the graph, and we sort them, and after we take the biggest as the ECG.
Alpha centrality (M)	AC_i	The alpha centrality measure, by vertex, can be considered as a generalization of eigenvector centrality (Bonacich e Lloyd (2001) <i>apud</i> Csardi e Nepusz (2006)).	AC is the solution of the equation $x = aA^T x + e$, where e is the vector of exogenous sources of status of the vertices, and a is the relative importance of the endogenous versus exogenous factors. In our case it is used $e = 1$ and $a = 1$.
Bonacich power centrality (M)	BPC_i	BPC_i measure corresponds to the notion that the power of a vertex is recursively defined by the sum of the power of its alters (Bonacich (1972), Bonacich (1987) <i>apud</i> Csardi e Nepusz (2006)).	$BPC = a(\mathbf{I} - bA)^{-1}A\mathbf{1}$, where a is a scaling parameter, b is an attenuation parameter and $a^2 + b^2 = n = V $. A is the adjacency matrix, \mathbf{I} is the identity matrix, and $\mathbf{1}$ is a column vector of ones.
Cocitation coupling (M)	CC_{ij}	Two vertices are cocited if there is another vertex citing both of them.	CC_{ij} simply counts how many times two vertices i and j are cocited (CSARDI; NEPUSZ, 2006).
Burt's constraint (M)	BC_i	Essentially BC_i is a measure of the extent to which ego is invested in people who are invested in other of ego's alters.	$BC_i = \sum_{i \neq j} (p_{ij} + \sum_{q \neq \{i,j\}} p_{iq} p_{qj})^2$; $i, j, q \in V$, with $p_{ij} = [A(i, j) + A(j, i)] / [\sum_{k \neq i} A(i, k) + A(k, i)]$, where A is the adjacency matrix (Burt (2004) <i>apud</i> Csardi e Nepusz (2006)).
Graph diversity (M)	GD_i	GD_i is a measure of diversity for all vertices (Eagle, Macy e Claxton (2010) <i>apud</i> Csardi e Nepusz (2006)).	$GD_i = \frac{H(i)}{\log \deg(i)}$, with $H(i) = - \sum_{j=1}^{\deg(i)} p_{ij} \log(p_{ij})$ and $p_{ij} = w_{ij} / \sum_{l=1}^{\deg(i)} w_{il}$, where w_{ij} is the weight of the edges. In our work, the weight of edges is considered unitary.
Kleinberg's hub centrality (M)	KHC_i	The hub scores of the vertices are defined as the coordinates of the principal eigenvector of AA^T , where A is the adjacency matrix of the graph (Kleinberg (1999) <i>apud</i> Csardi e Nepusz (2006)).	Compute AA^T , and take its eigenvalues. The eigenvector associate to the biggest eigenvalue is the KHC_i vector.
Kleinberg's hub centrality global (I)	KHCG	KHCG is the corresponding eigenvalue of the computed principal eigenvector KHC_i (Kleinberg (1999) <i>apud</i> Csardi e Nepusz (2006)).	KHCG is the biggest eigenvalue of the matrix AA^T .
Page rank (M)	PR_i	PR_i is the Google PageRank for the vertices.	PR_i is computed by the algorithm from Brin e Page (1998) <i>apud</i> Csardi e Nepusz (2006).

Continues on the next page...

...continued from previous page.			
Page rank global (I)	PRG	PRG is the eigenvalue corresponding to the eigenvector with the page rank scores.	PRG _i is also computed by the algorithm from Brin e Page (1998) <i>apud</i> Csardi e Nepusz (2006).
Jaccard similarity (M)	JS _{ij}	JS _{ij} is a similarity score for vertices based on their connection patterns. The Jaccard similarity coefficient of two vertices <i>i</i> and <i>j</i> is the number of common neighbors divided by the number of vertices that are neighbors of at least one of <i>i</i> and <i>j</i> (Adamic e Adar (2003) <i>apud</i> Csardi e Nepusz (2006)).	(description is enough for computation)
Dice similarity (M)	DS _{ij}	DS _{ij} is a similarity score for vertices based on their connection patterns. The dice similarity coefficient of two vertices <i>i</i> and <i>j</i> is twice the number of common neighbors divided by the sum of the degrees of <i>i</i> and <i>j</i> (Adamic e Adar (2003) <i>apud</i> Csardi e Nepusz (2006)).	(description is enough for computation)
Inverse log-weighted similarity (M)	ILWS _{ij}	ILWS _{ij} is a similarity score for vertices based on their connection patterns. It is based on the assumption that two vertices <i>i</i> and <i>j</i> should be considered more similar if they share a low-degree common neighbor, since high-degree common neighbors are more likely to appear even by pure chance. The inverse log-weighted similarity of two vertices <i>i</i> and <i>j</i> is the number of neighbors common to <i>i</i> and <i>j</i> , weighted by the inverse logarithm of their degrees (Adamic e Adar (2003) <i>apud</i> Csardi e Nepusz (2006)).	(description is enough for computation)
Subgraph centrality (M)	SC _i	Subgraph centrality of a vertex <i>i</i> measures the number of subgraphs a vertex participates in, weighting them according to their order.	The subgraph centrality of a vertex is defined as the number of closed loops originating at the vertex, where longer loops are exponentially downweighted. Currently the computation is performed by explicitly computing of all eigenvalues and eigenvectors of the adjacency matrix of the graph (Estrada e Rodriguez-Velazquez (2005) <i>apud</i> Csardi e Nepusz (2006)).

APÊNDICE B – Valores de $I(\lambda; k)$ para todo invariante k , com todas as redes juntas

SD = Standard Deviation.

CV = Coefficient of Variation.

Invariants	k	I($\lambda; k$)			
		Minimum	Mean	Median	Maximum
Maximum of edge betweenness	1	4.92	5.00	5.00	5.08
SD of edge betweenness	2	3.57	3.64	3.65	3.70
Amplitude of edge betweenness	3	3.47	3.54	3.54	3.61
Kirchhoff index	4	3.18	3.25	3.25	3.32
Mean of edge betweenness	5	3.15	3.21	3.21	3.28
Mean of vertex betweenness	6	3.02	3.07	3.07	3.13
Mean distance	7	2.75	2.80	2.80	2.84
Mean transmission	8	2.72	2.76	2.76	2.81
Maximum transmission	9	2.70	2.75	2.75	2.80
Wiener index	10	2.64	2.68	2.68	2.73
Median transmission	11	2.63	2.68	2.68	2.73
CV of adjusted communicability distance	12	2.57	2.62	2.62	2.67
Median of edge betweenness	13	2.53	2.57	2.57	2.62
Harary index	14	2.47	2.52	2.52	2.58
SD of transmission	15	2.31	2.35	2.35	2.40
Mean eccentricity	16	2.30	2.34	2.34	2.37
Amplitude of transmission	17	2.30	2.34	2.34	2.40
CV of modularity	18	2.26	2.30	2.31	2.34
Maximum of adjusted betweenness centrality	19	2.26	2.29	2.29	2.34
Minimum transmission	20	2.25	2.29	2.29	2.33
SD of vertex betweenness	21	2.19	2.22	2.22	2.26
SD of adjusted betweenness centrality	22	2.19	2.22	2.22	2.26
Maximum of vertex betweenness	23	2.19	2.22	2.22	2.26
Vertex betweenness centrality	24	2.15	2.18	2.18	2.22
Mean of adjusted communicability distance	25	2.15	2.18	2.18	2.20
Amplitude of vertex betweenness	26	2.12	2.15	2.15	2.19
Amplitude of adjusted betweenness centrality	27	2.12	2.15	2.15	2.19
Vertices number	28	2.10	2.14	2.14	2.17
Interquartile Distance of edge betweenness	29	2.06	2.08	2.08	2.11
SD by median of edge betweenness	30	2.01	2.04	2.04	2.07
Median of adjusted betweenness centrality	31	1.95	1.99	1.99	2.03
SD of adjusted communicability distance	32	1.93	1.95	1.95	1.96
CV of cocitation couplin	33	1.86	1.90	1.90	1.92
CV of dice similarity	34	1.86	1.89	1.89	1.92
CV of inverse log-weighted similarity	35	1.83	1.86	1.86	1.89
Mean of inverse log-weighted similarity	36	1.84	1.86	1.86	1.88
Minimum of adjusted betweenness centrality	37	1.82	1.84	1.84	1.87
Interquartile Distance of vertex betweenness	38	1.76	1.80	1.80	1.82
Interquartile Distance of adjusted betweenness centrality	39	1.76	1.80	1.80	1.82
Algebraic connectivity	40	1.75	1.78	1.78	1.81
CV of Jaccard similarity	41	1.72	1.75	1.75	1.77
Median of vertex betweenness	42	1.71	1.74	1.74	1.77
SD by median of vertex betweenness	43	1.66	1.69	1.69	1.72
SD by median of adjusted betweenness centrality	44	1.66	1.69	1.69	1.72
SD of distance	45	1.63	1.66	1.66	1.67
Mean of cocitation coupling	46	1.63	1.66	1.65	1.68
CV of adjacent triangles	47	1.61	1.65	1.65	1.68
Interquartile Distance of transmission	48	1.61	1.65	1.65	1.68
Interquartile Distance of adjusted communicability distance	49	1.63	1.64	1.65	1.66
Interquartile Distance of modularity	50	1.60	1.62	1.62	1.63
Diameter	51	1.57	1.60	1.60	1.63
Edges number	52	1.54	1.56	1.56	1.58
Degree centralization	53	1.52	1.55	1.55	1.57
SD by median of transmission	54	1.52	1.54	1.54	1.56
Median eccentricity	55	1.48	1.50	1.50	1.51
Median of inverse log-weighted similarity	56	1.48	1.50	1.50	1.51
CV by median of dice similarity	57	1.46	1.49	1.49	1.51

Invariants	k	I(λ ;k)			
		Minimum	Mean	Median	Maximum
Median of dice similarity	58	1.43	1.44	1.44	1.46
Number of minimum sets that disconnects the graph	59	1.39	1.41	1.41	1.43
CV of eccentricity	60	1.38	1.40	1.40	1.43
CV by median of Jaccard similarity	61	1.38	1.40	1.40	1.42
Minimum eccentricity	62	1.38	1.40	1.40	1.41
CV by median of inverse log-weighted similarity	63	1.36	1.39	1.39	1.40
SD by median of inverse log-weighted similarity	64	1.36	1.37	1.37	1.39
Edge density	65	1.35	1.36	1.36	1.36
Mean of vertex degree normalized	66	1.35	1.36	1.36	1.36
Median of Jaccard similarity	67	1.34	1.35	1.35	1.36
CV by median of cocitation coupling	68	1.32	1.34	1.34	1.35
Median of cocitation coupling	69	1.32	1.33	1.33	1.35
CV of eigenvector centrality by vertices	70	1.31	1.32	1.32	1.34
CV of Kleinberg's hub centrality	71	1.31	1.32	1.32	1.34
Maximum of vertex degree normalized	72	1.31	1.32	1.32	1.34
Median of vertex degree normalized	73	1.31	1.32	1.32	1.33
Interquartilic distance of eccentricity	74	1.28	1.30	1.29	1.31
Minimum of eigenvector centrality by vertices	75	1.28	1.29	1.29	1.30
Amplitude of eigenvector centrality by vertices	76	1.28	1.29	1.29	1.30
Minimum of Kleinberg's hub centrality	77	1.28	1.29	1.29	1.30
Amplitude of Kleinberg's hub centrality	78	1.28	1.29	1.29	1.30
SD by median of dice similarity	79	1.27	1.28	1.28	1.30
Maximum of closeness centrality normalized	80	1.26	1.27	1.27	1.28
Global modularity betweenness	81	1.26	1.27	1.27	1.28
Minimum of edge betweenness	82	1.23	1.24	1.24	1.25
Communication efficiency	83	1.23	1.24	1.24	1.25
Median of closeness centrality normalized	84	1.23	1.23	1.23	1.24
Mean of closeness centrality normalized	85	1.22	1.23	1.23	1.24
Interquartilic distance of inverse log-weighted similarity	86	1.21	1.22	1.22	1.24
Mean of dice similarity	87	1.21	1.22	1.22	1.23
Mean of eigenvector centrality by vertices	88	1.20	1.21	1.21	1.22
Mean of Kleinberg's hub centrality	89	1.20	1.21	1.21	1.22
CV of local transitivity	90	1.18	1.21	1.21	1.24
Minimum of closeness centrality normalized	91	1.20	1.21	1.21	1.21
SD by median of Jaccard similarity	92	1.19	1.20	1.20	1.21
CV of edge betweenness	93	1.18	1.19	1.19	1.20
SD by median of cocitation coupling	94	1.17	1.18	1.18	1.19
Closeness centralization	95	1.17	1.18	1.18	1.19
Minimum of subgraph centrality	96	1.17	1.18	1.18	1.19
CV by median of modularity	97	1.16	1.18	1.18	1.19
Median of eigenvector centrality by vertices	98	1.17	1.17	1.17	1.18
Median of Kleinberg's hub centrality	99	1.17	1.17	1.17	1.18
CV by median of Bonacich power centrality	100	1.14	1.17	1.17	1.21
Minimum of average nearest neighbor degree	101	1.16	1.17	1.17	1.18
Proportion of largest clique	102	1.16	1.16	1.16	1.16
SD by median of modularity	103	1.15	1.16	1.16	1.17
CV by median of distance	104	1.14	1.15	1.15	1.17
CV by median of eigenvector centrality by vertices	105	1.13	1.14	1.14	1.15
CV by median of Kleinberg's hub centrality	106	1.13	1.14	1.14	1.15
CV of adjusted betweenness centrality	107	1.12	1.13	1.13	1.15
SD of inverse log-weighted similarity	108	1.13	1.13	1.13	1.14
Number of spanning trees	109	1.12	1.13	1.13	1.14
Interquartilic distance of dice similarity	110	1.12	1.13	1.13	1.14
Amplitude of average nearest neighbor degree	111	1.11	1.12	1.12	1.14
SD of cocitation coupling	112	1.11	1.12	1.12	1.13
Median of local transitivity	113	1.10	1.12	1.12	1.13
Amplitude of vertex degree normalized	114	1.10	1.11	1.11	1.12
CV of distance	115	1.09	1.10	1.10	1.12
CV by median of adjacent triangles	116	1.09	1.10	1.10	1.11
Minimum modularity	117	1.09	1.09	1.09	1.10
CV of average nearest neighbor degree	118	1.08	1.09	1.09	1.10
Maximum modularity	119	1.09	1.09	1.09	1.10
SD of eccentricity	120	1.08	1.09	1.09	1.10
CV by median of adjusted betweenness centrality	121	1.08	1.09	1.09	1.10
SD by median of distance	122	1.08	1.09	1.09	1.10
Amplitude of eccentricity	123	1.08	1.09	1.09	1.10
CV of Bonacich power centrality	124	1.00	1.09	1.00	1.24
CV by median of Burt's constraint	125	1.08	1.09	1.09	1.10
CV by median of transmission	126	1.07	1.09	1.09	1.10
Maximum of Bonacich power centrality	127	1.07	1.08	1.08	1.09

Invariants	k	I(λ ;k)			
		Minimum	Mean	Median	Maximum
Amplitude of Bonacich power centrality	128	1.07	1.08	1.08	1.09
Global transitivity	129	1.07	1.08	1.08	1.09
CV by median of edge betweenness	130	1.07	1.08	1.07	1.09
Median of adjacent triangles	131	1.07	1.07	1.07	1.08
CV by median of local transitivity	132	1.06	1.07	1.07	1.09
Median of subgraph centrality	133	1.06	1.07	1.07	1.08
Natural connectivity	134	1.06	1.07	1.07	1.08
CV by median of eccentricity	135	1.06	1.07	1.07	1.07
SD by median of eccentricity	136	1.06	1.07	1.07	1.07
Mean of adjacent triangles	137	1.05	1.06	1.06	1.07
CV by median of closeness centrality	138	1.05	1.06	1.06	1.07
CV by median of closeness centrality normalized	139	1.05	1.06	1.06	1.07
Mean of Jaccard similarity	140	1.06	1.06	1.06	1.07
Mean of local transitivity	141	1.05	1.06	1.06	1.07
Minimum of vertex degree normalized	142	1.06	1.06	1.06	1.06
Mean of subgraph centrality	143	1.05	1.06	1.06	1.06
SD of average nearest neighbor degree	144	1.05	1.05	1.05	1.06
Amplitude of Burt's constraint	145	1.05	1.05	1.05	1.06
Median distance	146	1.05	1.05	1.05	1.06
Mean of Bonacich power centrality	147	1.04	1.05	1.05	1.06
Minimum of Burt's constraint	148	1.04	1.05	1.05	1.05
Minimum of Bonacich power centrality	149	1.04	1.05	1.05	1.06
Assortativity degree	150	1.04	1.05	1.05	1.06
CV by median of average nearest neighbor degree	151	1.04	1.05	1.05	1.05
Ratio vertices with minimum degree	152	1.04	1.05	1.05	1.05
SD by median of adjacent triangles	153	1.04	1.04	1.04	1.05
CV of vertex degree	154	1.03	1.04	1.04	1.05
CV of vertex degree normalized	155	1.03	1.04	1.04	1.05
Eigenvector centrality global	156	1.03	1.04	1.04	1.04
SD of Bonacich power centrality	157	1.03	1.04	1.04	1.04
Interquartile distance of Bonacich power centrality	158	1.03	1.04	1.04	1.04
Interquartile distance of vertex degree	159	1.03	1.03	1.03	1.04
Kleinberg's hub centrality global	160	1.03	1.03	1.03	1.04
Interquartile distance of average nearest neighbor degree	161	1.03	1.03	1.03	1.04
Median of average nearest neighbor degree	162	1.02	1.03	1.03	1.04
SD by median of local transitivity	163	1.03	1.03	1.03	1.04
Amplitude of vertex degree	164	1.02	1.03	1.03	1.04
Mean of average nearest neighbor degree	165	1.02	1.03	1.03	1.04
Maximum of vertex degree	166	1.02	1.03	1.03	1.04
Mean of vertex degree	167	1.02	1.03	1.03	1.04
CV by median of vertex betweenness	168	1.00	1.03	1.00	1.14
CV by median of vertex betweenness normalized	169	1.00	1.03	1.00	1.14
SD by median of average nearest neighbor degree	170	1.02	1.03	1.03	1.03
CV of Burt's constraint	171	1.02	1.03	1.03	1.04
Median of Burt's constraint	172	1.02	1.02	1.02	1.03
CV of subgraph centrality	173	1.02	1.02	1.02	1.03
CV of transmission	174	1.02	1.02	1.02	1.03
Median of Bonacich power centrality	175	1.01	1.02	1.02	1.03
Maximum of cocitation coupling	176	1.01	1.02	1.02	1.03
Amplitude of cocitation coupling	177	1.01	1.02	1.02	1.03
SD of adjacent triangles	178	1.01	1.02	1.02	1.03
Median modularity	179	1.02	1.02	1.02	1.02
Interquartile distance of eigenvector centrality by vertices	180	1.02	1.02	1.02	1.02
Interquartile distance of Kleinberg's hub centrality	181	1.02	1.02	1.02	1.02
Maximum of Burt's constraint	182	1.01	1.02	1.02	1.02
Interquartile distance of Jaccard similarity	183	1.01	1.02	1.02	1.02
Maximum of average nearest neighbor degree	184	1.01	1.02	1.02	1.03
SD by median of Bonacich power centrality	185	1.01	1.02	1.02	1.02
Maximum of adjacent triangles	186	1.01	1.02	1.02	1.02
Interquartile distance of local transitivity	187	1.01	1.01	1.01	1.02
Ratio vertices with maximum degree	188	1.01	1.01	1.01	1.02
Maximum of leverage centrality	189	1.01	1.01	1.01	1.02
Maximum of Jaccard similarity	190	1.01	1.01	1.01	1.02
Amplitude of Jaccard similarity	191	1.01	1.01	1.01	1.02
Maximum of dice similarity	192	1.01	1.01	1.01	1.02
Amplitude of dice similarity	193	1.01	1.01	1.01	1.02
Amplitude of adjacent triangles	194	1.01	1.01	1.01	1.02
Maximum of alpha centrality	195	1.01	1.01	1.01	1.02
Amplitude of alpha centrality	196	1.01	1.01	1.01	1.02
Amplitude of vertex betweenness normalized	197	1.01	1.01	1.01	1.02

Invariants	k	I(λ ;k)			
		Minimum	Mean	Median	Maximum
CV of closeness centrality	198	1.01	1.01	1.01	1.02
CV of closeness centrality normalized	199	1.01	1.01	1.01	1.02
Interquartile distance of vertex degree normalized	200	1.01	1.01	1.01	1.02
Normalized degree centralization	201	1.01	1.01	1.01	1.02
Maximum of vertex betweenness normalized	202	1.01	1.01	1.01	1.02
Vertex betweenness centrality normalized	203	1.01	1.01	1.01	1.02
CV by median of page rank	204	1.00	1.01	1.01	1.02
Amount of largest cliques	205	1.01	1.01	1.01	1.01
SD of vertex degree normalized	206	1.01	1.01	1.01	1.01
SD of eigenvector centrality by vertices	207	1.01	1.01	1.01	1.01
Interquartile distance of Burt's constraint	208	1.01	1.01	1.01	1.01
SD of Kleinberg's hub centrality	209	1.01	1.01	1.01	1.01
Amplitude of leverage centrality	210	1.00	1.01	1.01	1.01
SD of vertex degree	211	1.00	1.01	1.01	1.02
Sd by median of eigenvector centrality by vertices	212	1.00	1.01	1.01	1.01
Sd by median of Kleinberg's hub centrality	213	1.00	1.01	1.01	1.01
CV of alpha centrality	214	1.00	1.01	1.01	1.01
Maximum of subgraph centrality	215	1.00	1.01	1.01	1.01
Interquartile distance of cocitation coupling	216	1.00	1.00	1.00	1.01
CV by median of leverage centrality	217	0.97	1.00	1.00	1.04
CV of page rank	218	1.00	1.00	1.00	1.01
Amplitude of subgraph centrality	219	1.00	1.00	1.00	1.01
SD of local transitivity	220	1.00	1.00	1.00	1.01
CV of leverage centrality	221	1.00	1.00	1.00	1.01
Maximum of inverse log-weighted similarity	222	0.99	1.00	1.00	1.01
Amplitude of inverse log-weighted similarity	223	0.99	1.00	1.00	1.01
Amplitude of modularity	224	1.00	1.00	1.00	1.01
SD by median of vertex degree	225	0.99	1.00	1.00	1.01
Mean of Burt's constraint	226	1.00	1.00	1.00	1.01
Edge connectivity	227	1.00	1.00	1.00	1.00
Vertex connectivity	228	1.00	1.00	1.00	1.00
Minimum of vertex degree	229	1.00	1.00	1.00	1.00
Minimum of adjusted communicability distance	230	1.00	1.00	1.00	1.00
Maximum of adjusted communicability distance	231	1.00	1.00	1.00	1.00
Median of adjusted communicability distance	232	1.00	1.00	1.00	1.00
Amplitude of adjusted communicability distance	233	1.00	1.00	1.00	1.00
SD by median of adjusted communicability distance	234	1.00	1.00	1.00	1.00
Minimum of local transitivity	235	1.00	1.00	1.00	1.00
Minimum of vertex betweenness normalized	236	1.00	1.00	1.00	1.00
Mean of vertex betweenness normalized	237	1.00	1.00	1.00	1.00
Median of vertex betweenness normalized	238	1.00	1.00	1.00	1.00
Interquartile distance of vertex betweenness normalized	239	1.00	1.00	1.00	1.00
SD by median of vertex betweenness normalized	240	1.00	1.00	1.00	1.00
Vertex betweenness normalized	241	1.00	1.00	1.00	1.00
Mean modulation	242	1.00	1.00	1.00	1.00
SD of modulation	243	1.00	1.00	1.00	1.00
Minimum of adjacent triangles	244	1.00	1.00	1.00	1.00
Minimum of alpha centrality	245	1.00	1.00	1.00	1.00
Minimum of leverage centrality	246	1.00	1.00	1.00	1.00
Mean of leverage centrality	247	1.00	1.00	1.00	1.00
Minimum of closeness centrality	248	1.00	1.00	1.00	1.00
Maximum of closeness centrality	249	1.00	1.00	1.00	1.00
Mean of closeness centrality	250	1.00	1.00	1.00	1.00
Median of closeness centrality	251	1.00	1.00	1.00	1.00
Amplitude of closeness centrality	252	1.00	1.00	1.00	1.00
Interquartile distance of closeness centrality	253	1.00	1.00	1.00	1.00
SD by median of closeness centrality	254	1.00	1.00	1.00	1.00
SD of closeness centrality	255	1.00	1.00	1.00	1.00
Amplitude of closeness centrality normalized	256	1.00	1.00	1.00	1.00
Interquartile distance of closeness centrality normalized	257	1.00	1.00	1.00	1.00
SD by median of closeness centrality normalized	258	1.00	1.00	1.00	1.00
SD of closeness centrality normalized	259	1.00	1.00	1.00	1.00
Maximum of eigenvector centrality by vertices	260	1.00	1.00	1.00	1.00
Minimum of cocitation coupling	261	1.00	1.00	1.00	1.00
SD by median of Burt's constraint	262	1.00	1.00	1.00	1.00
SD of Burt's constraint	263	1.00	1.00	1.00	1.00
Minimum of graph diversity	264	1.00	1.00	1.00	1.00
Maximum of graph diversity	265	1.00	1.00	1.00	1.00
Mean of graph diversity	266	1.00	1.00	1.00	1.00
Median of graph diversity	267	1.00	1.00	1.00	1.00

Invariants	k	I($\lambda; k$)			
		Minimum	Mean	Median	Maximum
Amplitude of graph diversity	268	1.00	1.00	1.00	1.00
Interquartile distance of graph diversity	269	1.00	1.00	1.00	1.00
SD by median of graph diversity	270	1.00	1.00	1.00	1.00
CV by median of graph diversity	271	1.00	1.00	1.00	1.00
SD of graph diversity	272	1.00	1.00	1.00	1.00
CV of graph diversity	273	1.00	1.00	1.00	1.00
Maximum of Kleinberg's hub centrality	274	1.00	1.00	1.00	1.00
Minimum of page rank	275	1.00	1.00	1.00	1.00
Maximum of page rank	276	1.00	1.00	1.00	1.00
Mean of page rank	277	1.00	1.00	1.00	1.00
Median of page rank	278	1.00	1.00	1.00	1.00
Amplitude of page rank	279	1.00	1.00	1.00	1.00
Interquartile distance of page rank	280	1.00	1.00	1.00	1.00
SD by median of page rank	281	1.00	1.00	1.00	1.00
SD of page rank	282	1.00	1.00	1.00	1.00
Page rank global	283	1.00	1.00	1.00	1.00
Minimum of Jaccard similarity	284	1.00	1.00	1.00	1.00
SD of Jaccard similarity	285	1.00	1.00	1.00	1.00
Minimum of dice similarity	286	1.00	1.00	1.00	1.00
SD of dice similarity	287	1.00	1.00	1.00	1.00
Minimum of inverse log-weighted similarity	288	1.00	1.00	1.00	1.00
SD of subgraph centrality	289	0.99	1.00	1.00	1.01
SD of leverage centrality	290	0.99	1.00	1.00	1.00
Closeness centralization normalized	291	0.99	1.00	1.00	1.00
SD by median of vertex degree normalized	292	0.99	1.00	1.00	1.00
Mean of alpha centrality	293	0.99	1.00	1.00	1.00
SD of alpha centrality	294	0.99	1.00	1.00	1.00
Interquartile distance of distance	295	0.99	1.00	1.00	1.00
CV by median of subgraph centrality	296	0.99	1.00	1.00	1.00
CV of vertex betweenness	297	0.99	1.00	1.00	1.00
CV of vertex betweenness normalized	298	0.99	1.00	1.00	1.00
Interquartile distance of adjacent triangles	299	0.99	1.00	1.00	1.01
Median of leverage centrality	300	0.99	0.99	0.99	1.00
Minimum of vertex betweenness	301	0.99	0.99	0.99	1.00
Amplitude of local transitivity	302	0.99	0.99	0.99	1.00
Median of vertex degree	303	0.99	0.99	0.99	1.00
Maximum of local transitivity	304	0.99	0.99	0.99	1.00
CV by median of vertex degree	305	0.98	0.99	0.99	1.00
CV by median of vertex degree normalized	306	0.98	0.99	0.99	1.00
Interquartile distance of alpha centrality	307	0.98	0.99	0.99	1.00
Interquartile distance of leverage centrality	308	0.99	0.99	0.99	0.99
SD by median of leverage centrality	309	0.99	0.99	0.99	0.99
SD by median of alpha centrality	310	0.98	0.99	0.99	1.00
Median of alpha centrality	311	0.98	0.99	0.99	1.00
Interquartile distance of subgraph centrality	312	0.98	0.99	0.99	0.99
SD by median of subgraph centrality	313	0.98	0.99	0.99	0.99
CV by median of alpha centrality	314	0.98	0.99	0.99	0.99
CV by median of adjusted communicability distance	315	NA	NA	NA	NA

NA = not available (due to the value of the invariant *per se* is an indeterminate value, with division by zero).

APÊNDICE C – Valores de $I(\lambda; k)$ para todo invariante k , com as redes separadas por n

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$
1	1	2.6397	1	2.6860	1	2.7058	1	2.7019	1	2.6739	1	2.6240	1	2.5672	1	2.5139	1	2.4863	1	2.4347	1	2.4067
2	40	2.1316	40	2.1655	40	2.1882	2	2.2270	2	2.2543	2	2.2566	2	2.2636	2	2.2649	2	2.2676	2	2.2588	2	2.2500
3	2	2.0029	2	2.1082	2	2.1600	40	2.2221	40	2.2310	40	2.2131	40	2.1902	40	2.1707	40	2.1494	40	2.1276	40	2.1058
4	4	1.9470	4	1.9846	4	1.9947	3	2.0123	3	2.0216	3	2.0277	3	2.0269	3	2.0213	3	2.0200	3	2.0069	3	2.0032
5	6	1.8976	3	1.9226	3	1.9628	4	1.9976	4	1.9880	4	1.9584	4	1.9520	4	1.9431	4	1.9434	4	1.9088	4	1.9205
6	10	1.8941	10	1.8756	10	1.8879	10	1.8786	10	1.8770	10	1.8414	10	1.8261	10	1.8106	10	1.8190	10	1.7753	10	1.7796
7	8	1.8941	8	1.8756	8	1.8879	8	1.8786	8	1.8770	8	1.8414	8	1.8261	8	1.8106	8	1.8190	8	1.7753	8	1.7796
8	3	1.8400	6	1.8517	6	1.8712	6	1.8757	6	1.8770	6	1.8414	6	1.8261	6	1.8106	6	1.8190	6	1.7753	6	1.7796
9	5	1.7972	5	1.8372	14	1.8179	14	1.7929	14	1.7740	14	1.7399	14	1.7209	14	1.6959	14	1.7191	14	1.6697	9	1.6746
10	14	1.7453	14	1.7420	5	1.7921	5	1.7374	5	1.7263	5	1.6754	5	1.6704	9	1.6466	9	1.6593	9	1.6570	14	1.6677
11	11	1.6107	7	1.5971	7	1.6193	7	1.6323	9	1.6155	9	1.6259	9	1.6379	5	1.6438	5	1.6569	5	1.6122	5	1.6188
12	7	1.5893	59	1.5813	11	1.6112	9	1.6040	11	1.6115	7	1.6139	7	1.6144	7	1.6070	7	1.6087	16	1.5901	7	1.6015
13	59	1.5855	53	1.5497	59	1.5843	126	1.5738	7	1.6067	16	1.5527	11	1.5905	16	1.5840	16	1.6043	7	1.5859	57	1.5879
14	126	1.5414	52	1.5417	9	1.5630	59	1.5651	59	1.5357	11	1.5516	16	1.5699	11	1.5451	11	1.5990	11	1.5364	63	1.5862
15	52	1.5207	167	1.5417	52	1.5197	11	1.5446	16	1.5284	59	1.5283	101	1.5152	101	1.4972	101	1.5177	57	1.5010	61	1.5861
16	167	1.5207	12	1.5417	167	1.5197	16	1.4936	101	1.4973	126	1.5279	59	1.5134	59	1.4853	93	1.4881	101	1.5003	68	1.5859
17	12	1.5207	11	1.5405	12	1.5197	52	1.4903	52	1.4752	101	1.4916	93	1.4758	93	1.4775	59	1.4793	63	1.4990	16	1.5802
18	9	1.5071	9	1.5208	126	1.5085	167	1.4903	167	1.4752	93	1.4548	52	1.4441	126	1.4762	63	1.4674	61	1.4983	11	1.5745
19	60	1.4827	126	1.4955	101	1.4925	12	1.4903	12	1.4752	52	1.4493	167	1.4441	45	1.4421	61	1.4673	68	1.4982	101	1.5288
20	96	1.4560	101	1.4811	16	1.4685	101	1.4658	93	1.4398	167	1.4493	12	1.4441	115	1.4365	68	1.4660	93	1.4896	93	1.4977
21	16	1.4448	115	1.4700	25	1.4179	93	1.4331	45	1.4395	12	1.4493	45	1.4383	60	1.4301	57	1.4636	59	1.4595	59	1.4644
22	101	1.4417	16	1.4449	115	1.4104	45	1.4268	115	1.4136	45	1.4408	115	1.4331	52	1.4282	45	1.4499	126	1.4568	115	1.4535
23	115	1.4246	25	1.4236	45	1.4100	115	1.4093	126	1.4129	115	1.4351	29	1.3957	167	1.4282	115	1.4480	115	1.4458	45	1.4479
24	109	1.4231	109	1.4166	96	1.4077	29	1.3937	29	1.3960	29	1.3898	35	1.3834	12	1.4282	52	1.4321	45	1.4442	35	1.4162
25	25	1.4020	13	1.4114	13	1.3923	96	1.3845	96	1.3851	34	1.3736	34	1.3796	29	1.3987	167	1.4321	52	1.4076	52	1.4128
26	13	1.3985	46	1.4014	93	1.3854	13	1.3789	53	1.3760	35	1.3736	53	1.3755	35	1.3915	12	1.4321	167	1.4076	12	1.4128
27	46	1.3954	45	1.3871	109	1.3842	30	1.3704	30	1.3749	96	1.3706	30	1.3739	34	1.3834	35	1.4083	12	1.4076	15	1.4082
28	36	1.3873	96	1.3850	53	1.3783	138	1.3685	34	1.3656	30	1.3671	33	1.3722	33	1.3787	29	1.4068	35	1.4024	29	1.4049
29	45	1.3757	36	1.3719	46	1.3767	139	1.3685	35	1.3623	33	1.3649	96	1.3672	30	1.3749	60	1.4003	29	1.3994	167	1.4025
30	165	1.3444	138	1.3685	29	1.3687	34	1.3529	13	1.3607	15	1.3390	55	1.3645	63	1.3730	34	1.3998	60	1.3978	60	1.4011
31	156	1.3339	139	1.3685	30	1.3473	35	1.3472	33	1.3555	13	1.3351	15	1.3553	61	1.3707	33	1.3934	15	1.3892	34	1.4000
32	160	1.3315	29	1.3619	295	1.3388	33	1.3437	50	1.3508	41	1.3302	41	1.3382	68	1.3703	53	1.3896	34	1.3887	33	1.3881
33	29	1.3267	93	1.3612	34	1.3339	46	1.3415	74	1.3349	46	1.2987	13	1.3284	57	1.3696	15	1.3854	33	1.3800	53	1.3844
34	53	1.3193	165	1.3425	33	1.3268	109	1.3223	46	1.3214	74	1.2908	63	1.3021	15	1.3672	30	1.3811	30	1.3760	30	1.3790
35	162	1.3181	156	1.3374	165	1.3251	41	1.3076	109	1.3203	55	1.2892	17	1.3009	96	1.3489	41	1.3590	41	1.3528	41	1.3652
36	93	1.3157	30	1.3357	35	1.3210	15	1.3009	41	1.3190	109	1.2855	68	1.2976	41	1.3442	96	1.3446	17	1.3325	17	1.3490

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$
37	33	1.3085	160	1.3342	36	1.3176	165	1.2968	15	1.3181	17	1.2853	60	1.2972	55	1.3209	13	1.3257	96	1.3284	96	1.3325
38	30	1.3040	162	1.3115	156	1.3033	295	1.2964	49	1.3153	165	1.2762	61	1.2971	13	1.3157	17	1.3247	53	1.3086	13	1.3064
39	71	1.2982	34	1.3071	160	1.2982	25	1.2864	25	1.2940	138	1.2739	126	1.2933	17	1.3109	55	1.3072	13	1.2998	51	1.2889
40	70	1.2976	33	1.3041	162	1.2962	36	1.2797	165	1.2928	139	1.2739	57	1.2902	53	1.2928	51	1.2753	51	1.2861	21	1.2615
41	34	1.2960	134	1.2971	41	1.2867	156	1.2703	17	1.2717	53	1.2657	46	1.2804	46	1.2679	165	1.2623	21	1.2585	22	1.2615
42	143	1.2939	51	1.2920	51	1.2848	162	1.2665	162	1.2656	57	1.2537	165	1.2658	165	1.2595	46	1.2603	22	1.2585	165	1.2499
43	134	1.2917	35	1.2903	134	1.2693	160	1.2632	55	1.2606	162	1.2482	109	1.2570	21	1.2500	21	1.2506	165	1.2482	46	1.2334
44	35	1.2763	143	1.2890	15	1.2628	50	1.2625	36	1.2539	61	1.2479	56	1.2539	22	1.2500	22	1.2506	46	1.2374	162	1.2212
45	133	1.2660	295	1.2707	71	1.2621	17	1.2599	60	1.2516	63	1.2466	69	1.2511	69	1.2459	162	1.2346	70	1.2178	71	1.2206
46	74	1.2547	133	1.2702	70	1.2613	60	1.2593	156	1.2514	70	1.2463	58	1.2471	138	1.2429	71	1.2325	71	1.2178	70	1.2205
47	41	1.2464	71	1.2663	133	1.2601	71	1.2579	71	1.2513	71	1.2463	64	1.2441	139	1.2429	70	1.2324	162	1.2176	174	1.2038
48	18	1.2326	70	1.2658	143	1.2575	70	1.2577	70	1.2511	68	1.2446	21	1.2439	51	1.2427	126	1.2202	55	1.2120	123	1.2006
49	15	1.2169	41	1.2615	60	1.2383	53	1.2506	160	1.2472	21	1.2353	22	1.2439	56	1.2405	69	1.2041	138	1.2113	104	1.1944
50	51	1.2102	60	1.2550	17	1.2251	134	1.2430	134	1.2286	22	1.2353	70	1.2388	58	1.2382	58	1.2040	139	1.2113	19	1.1938
51	138	1.2050	15	1.2388	18	1.2085	51	1.2360	22	1.2261	156	1.2297	71	1.2387	64	1.2370	109	1.2026	109	1.1958	23	1.1937
52	139	1.2050	18	1.2308	21	1.1982	143	1.2247	21	1.2260	60	1.2261	162	1.2378	70	1.2316	156	1.2020	23	1.1939	146	1.1918
53	204	1.1863	17	1.1986	22	1.1982	21	1.2216	133	1.2246	160	1.2244	74	1.2329	79	1.2316	79	1.2020	19	1.1939	133	1.1896
54	17	1.1780	104	1.1944	152	1.1974	22	1.2216	143	1.2151	56	1.2241	79	1.2322	162	1.2315	120	1.2010	152	1.1937	156	1.1879
55	77	1.1760	122	1.1940	104	1.1824	133	1.2216	51	1.1973	58	1.2186	67	1.2275	71	1.2315	133	1.1990	174	1.1928	26	1.1856
56	78	1.1760	21	1.1901	62	1.1803	74	1.2144	152	1.1957	69	1.2181	156	1.2186	109	1.2238	23	1.1968	123	1.1927	27	1.1856
57	75	1.1758	22	1.1901	122	1.1800	49	1.2059	57	1.1922	67	1.2176	160	1.2136	67	1.2238	19	1.1968	156	1.1875	160	1.1835
58	76	1.1758	152	1.1820	23	1.1788	152	1.1974	23	1.1902	49	1.2156	51	1.2127	25	1.2191	67	1.1963	26	1.1873	134	1.1798
59	152	1.1723	23	1.1759	19	1.1788	23	1.1966	19	1.1902	36	1.2135	92	1.2101	94	1.2143	160	1.1953	27	1.1873	198	1.1798
60	120	1.1643	19	1.1759	20	1.1739	19	1.1966	56	1.1894	64	1.2118	25	1.2083	92	1.2128	152	1.1948	160	1.1837	199	1.1798
61	123	1.1605	20	1.1678	77	1.1622	18	1.1868	68	1.1877	134	1.2100	133	1.2077	156	1.2048	56	1.1938	133	1.1793	143	1.1729
62	22	1.1593	77	1.1566	78	1.1622	62	1.1826	63	1.1871	133	1.2037	94	1.2058	120	1.2039	64	1.1922	120	1.1785	109	1.1682
63	21	1.1589	78	1.1566	75	1.1620	26	1.1784	58	1.1836	79	1.2018	134	1.2023	160	1.2004	134	1.1904	134	1.1776	152	1.1678
64	20	1.1573	75	1.1565	76	1.1620	27	1.1784	69	1.1830	152	1.1994	152	1.1999	23	1.1963	74	1.1904	198	1.1713	130	1.1667
65	103	1.1567	76	1.1565	305	1.1613	20	1.1720	67	1.1829	143	1.1987	23	1.1941	19	1.1963	92	1.1892	199	1.1713	55	1.1604
66	23	1.1559	26	1.1552	306	1.1613	75	1.1649	61	1.1815	25	1.1984	19	1.1941	152	1.1950	174	1.1889	143	1.1704	20	1.1583
67	19	1.1559	27	1.1552	26	1.1588	76	1.1649	18	1.1775	51	1.1952	120	1.1939	134	1.1910	26	1.1876	25	1.1640	120	1.1557
68	97	1.1523	123	1.1547	27	1.1588	77	1.1649	64	1.1770	92	1.1948	36	1.1923	26	1.1851	27	1.1876	74	1.1637	126	1.1521
69	90	1.1420	74	1.1516	50	1.1569	78	1.1649	26	1.1752	23	1.1933	143	1.1915	27	1.1851	94	1.1839	130	1.1606	18	1.1487
70	305	1.1406	204	1.1440	138	1.1347	56	1.1593	27	1.1752	19	1.1933	26	1.1836	133	1.1846	143	1.1794	20	1.1541	74	1.1447
71	306	1.1406	62	1.1418	139	1.1347	55	1.1540	20	1.1716	26	1.1805	27	1.1836	143	1.1815	198	1.1680	18	1.1438	75	1.1383
72	303	1.1373	120	1.1398	303	1.1336	58	1.1533	79	1.1679	27	1.1805	174	1.1698	174	1.1751	199	1.1680	75	1.1415	76	1.1383
73	26	1.1340	81	1.1333	130	1.1307	69	1.1527	77	1.1648	18	1.1657	20	1.1621	36	1.1670	123	1.1659	76	1.1415	77	1.1383
74	27	1.1340	184	1.1289	55	1.1236	130	1.1506	78	1.1648	20	1.1649	75	1.1583	198	1.1579	130	1.1608	77	1.1415	78	1.1383
75	81	1.1329	215	1.1281	204	1.1216	104	1.1448	75	1.1647	120	1.1632	76	1.1583	199	1.1579	20	1.1585	78	1.1415	24	1.1308
76	184	1.1319	130	1.1271	79	1.1213	79	1.1414	76	1.1647	75	1.1617	77	1.1583	20	1.1571	25	1.1555	24	1.1312	48	1.1220
77	215	1.1299	90	1.1222	48	1.1212	122	1.1399	92	1.1532	76	1.1617	78	1.1583	130	1.1541	18	1.1538	104	1.1310	106	1.1189
78	295	1.1297	151	1.1216	56	1.1206	67	1.1340	130	1.1479	94	1.1617	18	1.1574	75	1.1521	36	1.1491	146	1.1256	105	1.1188
79	48	1.1209	97	1.1213	174	1.1190	64	1.1326	174	1.1457	77	1.1617	130	1.1550	76	1.1521	75	1.1483	36	1.1255	116	1.1184
80	104	1.1154	305	1.1155	81	1.1173	174	1.1301	198	1.1359	78	1.1617	198	1.1539	77	1.1520	76	1.1483	69	1.1239	25	1.1156

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$
81	122	1.1153	306	1.1155	74	1.1169	172	1.1298	199	1.1359	174	1.1583	199	1.1539	78	1.1520	77	1.1483	58	1.1239	54	1.1114
82	42	1.1149	303	1.1144	58	1.1167	198	1.1269	172	1.1222	50	1.1540	24	1.1254	18	1.1510	78	1.1483	79	1.1238	36	1.1109
83	31	1.1126	48	1.1081	69	1.1160	199	1.1269	94	1.1217	198	1.1476	172	1.1165	74	1.1444	24	1.1311	67	1.1195	107	1.1085
84	106	1.1120	38	1.1074	198	1.1115	204	1.1225	62	1.1215	199	1.1476	48	1.1154	24	1.1298	105	1.1184	56	1.1186	132	1.1071
85	105	1.1114	39	1.1074	199	1.1115	24	1.1203	120	1.1205	130	1.1468	106	1.1120	123	1.1228	106	1.1183	64	1.1180	38	1.1049
86	130	1.1112	202	1.1048	106	1.1106	68	1.1176	305	1.1194	172	1.1284	105	1.1118	172	1.1201	48	1.1163	92	1.1171	39	1.1049
87	176	1.1095	106	1.1046	184	1.1102	202	1.1093	306	1.1194	24	1.1230	38	1.1067	116	1.1174	54	1.1092	94	1.1157	118	1.1039
88	177	1.1095	42	1.1045	105	1.1100	197	1.1068	135	1.1187	48	1.1126	39	1.1067	106	1.1073	118	1.1044	116	1.1138	197	1.0931
89	73	1.1051	31	1.1045	215	1.1071	106	1.1057	24	1.1168	38	1.1110	54	1.1036	105	1.1071	172	1.1036	48	1.1128	202	1.0928
90	38	1.1045	79	1.1042	24	1.1055	105	1.1046	136	1.1165	39	1.1110	116	1.1030	107	1.1031	38	1.1034	132	1.1091	172	1.0871
91	39	1.1045	198	1.1039	172	1.1053	151	1.1044	48	1.1122	204	1.1069	202	1.1023	202	1.1018	39	1.1034	107	1.1088	125	1.0753
92	125	1.1027	199	1.1039	38	1.1047	81	1.1013	104	1.1117	105	1.1067	197	1.1019	197	1.1015	116	1.1031	105	1.1063	112	1.0748
93	110	1.1013	105	1.1037	39	1.1047	63	1.0992	106	1.1114	106	1.1065	107	1.1010	151	1.0998	107	1.1008	106	1.1063	43	1.0725
94	62	1.0998	24	1.1015	42	1.1046	38	1.0984	105	1.1109	202	1.1039	118	1.1004	118	1.0980	197	1.0987	38	1.1015	44	1.0725
95	44	1.0950	197	1.0997	31	1.1046	39	1.0984	38	1.1064	197	1.1032	305	1.0978	132	1.0975	202	1.0986	39	1.1015	32	1.0697
96	202	1.0940	125	1.0978	202	1.1036	116	1.0978	39	1.1064	107	1.1006	306	1.0978	48	1.0963	104	1.0960	151	1.1012	203	1.0685
97	198	1.0938	112	1.0969	67	1.1017	92	1.0951	202	1.1056	151	1.0992	303	1.0852	39	1.0942	132	1.0872	172	1.1012	81	1.0663
98	199	1.0938	219	1.0963	197	1.0989	107	1.0933	197	1.1040	116	1.0973	135	1.0848	38	1.0941	146	1.0837	118	1.0985	151	1.0616
99	79	1.0931	176	1.0950	90	1.0942	48	1.0919	295	1.1037	118	1.0954	32	1.0824	54	1.0938	32	1.0791	54	1.0978	67	1.0612
100	116	1.0909	177	1.0950	135	1.0923	305	1.0916	303	1.1027	54	1.0931	81	1.0822	104	1.0870	81	1.0762	197	1.0955	92	1.0608
101	219	1.0891	174	1.0945	97	1.0905	306	1.0916	122	1.1012	81	1.0919	136	1.0819	32	1.0866	112	1.0756	202	1.0954	31	1.0597
102	55	1.0880	172	1.0915	54	1.0899	32	1.0916	54	1.0971	305	1.0882	104	1.0765	204	1.0841	43	1.0751	125	1.0824	42	1.0596
103	197	1.0871	154	1.0909	136	1.0898	184	1.0900	81	1.0954	306	1.0882	123	1.0763	125	1.0810	44	1.0751	112	1.0712	79	1.0568
104	293	1.0859	155	1.0909	49	1.0896	42	1.0884	107	1.0910	104	1.0810	112	1.0758	81	1.0774	305	1.0735	81	1.0703	58	1.0567
105	166	1.0857	289	1.0898	112	1.0890	31	1.0884	204	1.0899	32	1.0799	43	1.0753	112	1.0737	306	1.0735	203	1.0702	303	1.0534
106	72	1.0857	116	1.0876	64	1.0870	94	1.0878	118	1.0889	184	1.0760	44	1.0753	203	1.0721	203	1.0718	32	1.0678	82	1.0523
107	112	1.0851	163	1.0862	44	1.0856	303	1.0852	116	1.0882	42	1.0760	203	1.0715	305	1.0702	303	1.0692	43	1.0652	69	1.0516
108	174	1.0824	293	1.0862	32	1.0807	120	1.0851	42	1.0856	31	1.0760	204	1.0701	306	1.0702	125	1.0678	44	1.0652	187	1.0512
109	43	1.0819	103	1.0841	118	1.0802	215	1.0848	31	1.0856	303	1.0756	42	1.0685	43	1.0679	42	1.0641	204	1.0652	293	1.0508
110	56	1.0815	57	1.0828	219	1.0802	118	1.0833	184	1.0844	112	1.0751	31	1.0685	44	1.0679	31	1.0641	305	1.0612	305	1.0504
111	137	1.0814	50	1.0818	123	1.0799	54	1.0815	32	1.0827	125	1.0732	132	1.0657	303	1.0638	151	1.0611	306	1.0612	306	1.0504
112	87	1.0805	113	1.0805	107	1.0797	163	1.0813	112	1.0799	132	1.0723	97	1.0638	184	1.0613	293	1.0585	42	1.0573	138	1.0499
113	24	1.0802	44	1.0805	176	1.0788	112	1.0807	138	1.0796	43	1.0720	184	1.0633	293	1.0588	184	1.0557	31	1.0573	139	1.0499
114	154	1.0790	56	1.0799	177	1.0788	125	1.0792	139	1.0796	44	1.0720	293	1.0620	146	1.0581	138	1.0546	303	1.0571	184	1.0495
115	155	1.0790	166	1.0781	43	1.0783	61	1.0791	215	1.0780	203	1.0718	90	1.0616	42	1.0571	139	1.0546	184	1.0541	94	1.0490
116	164	1.0786	107	1.0774	293	1.0769	90	1.0774	43	1.0744	215	1.0713	215	1.0606	31	1.0571	82	1.0539	293	1.0534	56	1.0487
117	114	1.0786	182	1.0764	125	1.0745	57	1.0762	44	1.0744	97	1.0707	125	1.0604	215	1.0560	90	1.0527	82	1.0529	64	1.0487
118	54	1.0780	137	1.0756	166	1.0737	113	1.0749	203	1.0702	293	1.0661	151	1.0599	82	1.0543	204	1.0508	88	1.0495	62	1.0479
119	113	1.0779	153	1.0753	289	1.0730	203	1.0732	293	1.0681	90	1.0647	50	1.0585	166	1.0539	89	1.0507	89	1.0495	88	1.0473
120	289	1.0769	58	1.0734	164	1.0698	43	1.0722	97	1.0675	122	1.0615	49	1.0569	164	1.0535	88	1.0506	215	1.0461	89	1.0473
121	150	1.0762	118	1.0733	80	1.0693	293	1.0708	166	1.0658	135	1.0615	138	1.0566	88	1.0535	166	1.0504	86	1.0458	121	1.0466
122	82	1.0756	43	1.0733	68	1.0691	44	1.0706	90	1.0655	136	1.0609	139	1.0566	89	1.0535	164	1.0503	166	1.0452	164	1.0453
123	80	1.0745	67	1.0724	182	1.0669	176	1.0690	164	1.0640	166	1.0596	82	1.0563	135	1.0523	215	1.0500	164	1.0451	86	1.0453
124	95	1.0743	32	1.0721	187	1.0668	177	1.0690	176	1.0597	164	1.0589	187	1.0560	187	1.0507	150	1.0476	150	1.0448	166	1.0451

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$
125	181	1.0741	69	1.0719	203	1.0640	166	1.0657	177	1.0597	163	1.0580	166	1.0556	136	1.0507	121	1.0469	187	1.0444	150	1.0448
126	180	1.0740	82	1.0712	211	1.0625	97	1.0649	219	1.0589	113	1.0570	164	1.0552	150	1.0484	182	1.0463	135	1.0438	215	1.0448
127	182	1.0712	294	1.0707	82	1.0622	219	1.0637	163	1.0571	89	1.0556	88	1.0535	182	1.0481	187	1.0461	182	1.0433	176	1.0433
128	294	1.0686	54	1.0706	137	1.0616	82	1.0635	125	1.0570	88	1.0555	89	1.0535	97	1.0478	176	1.0451	136	1.0415	177	1.0433
129	246	1.0683	217	1.0696	294	1.0613	164	1.0633	151	1.0564	62	1.0541	182	1.0493	90	1.0460	177	1.0451	176	1.0412	211	1.0432
130	163	1.0682	164	1.0679	154	1.0611	187	1.0632	82	1.0556	219	1.0538	150	1.0477	176	1.0453	211	1.0442	177	1.0412	182	1.0428
131	64	1.0668	180	1.0673	155	1.0611	153	1.0613	88	1.0555	150	1.0529	176	1.0470	177	1.0453	80	1.0423	121	1.0405	111	1.0412
132	84	1.0659	181	1.0673	181	1.0610	182	1.0606	89	1.0552	82	1.0522	177	1.0470	80	1.0427	111	1.0422	211	1.0397	295	1.0401
133	86	1.0652	73	1.0657	163	1.0607	289	1.0577	150	1.0548	80	1.0514	121	1.0466	219	1.0427	86	1.0410	294	1.0397	190	1.0398
134	118	1.0632	203	1.0641	180	1.0607	150	1.0569	182	1.0548	182	1.0507	219	1.0458	163	1.0425	180	1.0407	90	1.0394	191	1.0398
135	218	1.0619	84	1.0638	92	1.0596	89	1.0560	80	1.0547	176	1.0500	163	1.0452	121	1.0420	181	1.0407	98	1.0392	192	1.0391
136	211	1.0613	68	1.0618	113	1.0586	88	1.0558	289	1.0540	177	1.0500	294	1.0452	211	1.0410	294	1.0397	99	1.0392	193	1.0391
137	153	1.0612	150	1.0617	150	1.0572	80	1.0555	187	1.0536	187	1.0497	80	1.0451	98	1.0407	219	1.0394	111	1.0375	98	1.0388
138	91	1.0612	211	1.0607	89	1.0557	180	1.0541	181	1.0517	181	1.0497	289	1.0442	99	1.0406	98	1.0393	297	1.0375	99	1.0388
139	173	1.0605	64	1.0582	151	1.0556	181	1.0540	180	1.0515	289	1.0497	122	1.0440	113	1.0405	99	1.0393	298	1.0375	289	1.0375
140	89	1.0597	226	1.0571	226	1.0554	294	1.0531	294	1.0510	180	1.0496	180	1.0434	289	1.0396	135	1.0381	80	1.0367	297	1.0374
141	68	1.0594	246	1.0568	88	1.0553	137	1.0527	113	1.0504	294	1.0484	181	1.0434	294	1.0394	289	1.0377	289	1.0361	298	1.0374
142	88	1.0592	173	1.0567	153	1.0527	226	1.0525	153	1.0468	121	1.0426	211	1.0424	111	1.0392	221	1.0366	190	1.0345	90	1.0365
143	107	1.0589	187	1.0562	86	1.0519	135	1.0520	226	1.0467	153	1.0423	111	1.0418	180	1.0361	190	1.0360	191	1.0345	80	1.0362
144	159	1.0588	80	1.0560	120	1.0510	136	1.0516	154	1.0455	226	1.0415	113	1.0411	181	1.0361	191	1.0360	219	1.0345	204	1.0362
145	221	1.0572	89	1.0551	84	1.0505	132	1.0493	155	1.0455	211	1.0403	99	1.0401	221	1.0358	192	1.0355	180	1.0344	135	1.0359
146	172	1.0572	88	1.0547	94	1.0499	84	1.0479	132	1.0438	111	1.0403	98	1.0400	190	1.0356	193	1.0355	181	1.0344	219	1.0349
147	58	1.0557	195	1.0539	57	1.0499	154	1.0447	211	1.0429	99	1.0402	217	1.0384	191	1.0356	217	1.0350	192	1.0341	294	1.0345
148	186	1.0549	196	1.0539	246	1.0493	155	1.0447	137	1.0420	98	1.0401	131	1.0370	295	1.0354	136	1.0327	193	1.0341	180	1.0343
149	67	1.0535	186	1.0537	116	1.0476	86	1.0435	121	1.0415	221	1.0380	226	1.0365	192	1.0352	297	1.0324	62	1.0340	181	1.0343
150	226	1.0529	145	1.0514	195	1.0457	131	1.0424	111	1.0402	154	1.0379	221	1.0362	193	1.0352	298	1.0324	221	1.0327	221	1.0338
151	32	1.0525	194	1.0511	196	1.0457	121	1.0421	99	1.0396	155	1.0379	146	1.0353	131	1.0342	145	1.0319	295	1.0319	145	1.0314
152	195	1.0522	178	1.0487	173	1.0457	145	1.0417	98	1.0395	131	1.0377	153	1.0350	122	1.0336	163	1.0290	145	1.0308	173	1.0300
153	196	1.0522	87	1.0485	121	1.0442	201	1.0404	117	1.0388	307	1.0358	190	1.0346	145	1.0325	131	1.0286	307	1.0294	136	1.0290
154	203	1.0515	312	1.0479	145	1.0435	99	1.0403	131	1.0384	137	1.0345	191	1.0346	226	1.0318	113	1.0284	72	1.0260	154	1.0276
155	121	1.0510	313	1.0476	221	1.0412	98	1.0402	221	1.0373	123	1.0343	192	1.0339	297	1.0310	72	1.0276	114	1.0259	155	1.0276
156	69	1.0509	225	1.0472	205	1.0408	246	1.0396	145	1.0362	173	1.0343	193	1.0339	298	1.0310	144	1.0276	144	1.0259	144	1.0272
157	187	1.0507	292	1.0472	186	1.0407	173	1.0394	72	1.0354	145	1.0339	145	1.0333	153	1.0305	114	1.0275	173	1.0254	114	1.0266
158	131	1.0505	86	1.0469	99	1.0407	91	1.0391	195	1.0353	225	1.0328	137	1.0323	173	1.0296	226	1.0274	171	1.0249	72	1.0263
159	119	1.0497	55	1.0466	91	1.0406	225	1.0378	196	1.0353	117	1.0328	195	1.0307	72	1.0294	173	1.0269	154	1.0238	171	1.0244
160	151	1.0494	307	1.0456	98	1.0404	129	1.0373	86	1.0351	72	1.0326	196	1.0307	114	1.0291	137	1.0252	155	1.0238	168	1.0238
161	145	1.0488	205	1.0440	307	1.0403	195	1.0367	246	1.0350	190	1.0323	72	1.0300	154	1.0280	141	1.0246	163	1.0235	169	1.0238
162	194	1.0483	91	1.0428	63	1.0393	196	1.0367	173	1.0344	191	1.0323	297	1.0296	155	1.0280	295	1.0242	113	1.0226	95	1.0238
163	205	1.0481	129	1.0427	72	1.0390	141	1.0361	114	1.0342	192	1.0321	298	1.0296	137	1.0266	307	1.0231	95	1.0220	137	1.0235
164	57	1.0472	95	1.0427	194	1.0386	72	1.0358	84	1.0332	193	1.0321	154	1.0295	246	1.0252	195	1.0229	226	1.0220	141	1.0216
165	85	1.0470	221	1.0425	131	1.0382	211	1.0357	190	1.0319	195	1.0320	114	1.0295	144	1.0250	196	1.0229	141	1.0214	307	1.0210
166	129	1.0467	117	1.0415	128	1.0376	221	1.0344	191	1.0319	196	1.0320	155	1.0295	117	1.0246	246	1.0226	131	1.0213	220	1.0209
167	50	1.0446	72	1.0412	313	1.0370	111	1.0342	91	1.0317	114	1.0319	173	1.0289	141	1.0245	129	1.0222	246	1.0212	157	1.0209
168	220	1.0441	131	1.0412	129	1.0366	114	1.0339	192	1.0315	246	1.0301	307	1.0278	195	1.0243	95	1.0221	137	1.0211	246	1.0201

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$
169	212	1.0426	99	1.0409	103	1.0366	190	1.0334	193	1.0315	91	1.0300	141	1.0274	196	1.0243	313	1.0217	195	1.0208	226	1.0201
170	213	1.0424	128	1.0404	114	1.0365	191	1.0334	129	1.0305	84	1.0282	91	1.0274	307	1.0227	168	1.0216	196	1.0208	195	1.0192
171	65	1.0420	98	1.0404	111	1.0365	192	1.0331	159	1.0299	297	1.0280	129	1.0260	129	1.0224	169	1.0216	168	1.0204	196	1.0192
172	66	1.0420	121	1.0400	212	1.0352	193	1.0331	141	1.0294	298	1.0280	117	1.0259	225	1.0222	220	1.0212	169	1.0204	129	1.0182
173	99	1.0412	220	1.0390	213	1.0352	213	1.0328	205	1.0291	141	1.0266	144	1.0258	220	1.0222	171	1.0211	220	1.0200	313	1.0169
174	98	1.0411	92	1.0389	127	1.0345	212	1.0327	307	1.0286	212	1.0266	246	1.0256	91	1.0218	62	1.0209	129	1.0195	205	1.0167
175	141	1.0389	85	1.0385	117	1.0344	205	1.0324	212	1.0285	213	1.0266	220	1.0236	171	1.0218	153	1.0206	214	1.0192	214	1.0162
176	209	1.0377	213	1.0379	157	1.0335	186	1.0318	213	1.0285	129	1.0259	213	1.0233	95	1.0217	91	1.0196	157	1.0189	163	1.0157
177	149	1.0372	212	1.0377	85	1.0330	194	1.0309	313	1.0250	144	1.0257	212	1.0231	86	1.0211	212	1.0196	91	1.0189	212	1.0156
178	207	1.0372	224	1.0374	312	1.0315	85	1.0279	220	1.0243	86	1.0239	205	1.0220	212	1.0204	213	1.0196	212	1.0161	213	1.0155
179	178	1.0363	141	1.0347	220	1.0309	220	1.0276	186	1.0238	205	1.0229	214	1.0214	213	1.0204	154	1.0195	213	1.0161	117	1.0152
180	128	1.0354	159	1.0340	147	1.0309	117	1.0265	194	1.0237	220	1.0229	95	1.0213	157	1.0193	155	1.0195	117	1.0157	113	1.0148
181	225	1.0342	111	1.0333	141	1.0308	95	1.0261	297	1.0231	214	1.0221	84	1.0211	84	1.0184	117	1.0189	225	1.0156	224	1.0147
182	117	1.0342	214	1.0333	209	1.0302	128	1.0260	298	1.0231	313	1.0220	62	1.0209	62	1.0178	157	1.0186	224	1.0153	131	1.0144
183	94	1.0328	157	1.0329	207	1.0301	207	1.0260	85	1.0225	95	1.0211	168	1.0209	214	1.0173	214	1.0178	313	1.0137	178	1.0141
184	111	1.0316	114	1.0328	225	1.0293	209	1.0260	209	1.0224	207	1.0200	169	1.0209	168	1.0171	97	1.0171	205	1.0135	291	1.0141
185	208	1.0291	149	1.0324	159	1.0290	157	1.0250	144	1.0223	209	1.0200	313	1.0203	169	1.0171	205	1.0166	185	1.0135	122	1.0139
186	313	1.0291	207	1.0323	87	1.0286	214	1.0231	217	1.0223	312	1.0200	171	1.0198	205	1.0170	224	1.0161	153	1.0133	128	1.0135
187	214	1.0286	209	1.0322	201	1.0282	178	1.0229	207	1.0222	85	1.0194	207	1.0182	224	1.0159	128	1.0152	291	1.0132	91	1.0133
188	307	1.0282	218	1.0322	178	1.0282	123	1.0225	214	1.0216	157	1.0189	209	1.0182	186	1.0140	84	1.0129	178	1.0119	308	1.0125
189	92	1.0276	110	1.0322	214	1.0282	307	1.0217	95	1.0210	171	1.0189	157	1.0177	313	1.0139	161	1.0125	308	1.0118	161	1.0122
190	192	1.0272	127	1.0315	95	1.0276	127	1.0205	157	1.0208	186	1.0186	186	1.0161	194	1.0138	291	1.0122	128	1.0117	186	1.0119
191	191	1.0271	147	1.0287	190	1.0272	297	1.0197	128	1.0197	194	1.0185	194	1.0161	128	1.0138	308	1.0119	312	1.0111	194	1.0119
192	190	1.0270	94	1.0276	191	1.0272	298	1.0197	73	1.0187	224	1.0178	85	1.0161	308	1.0134	312	1.0112	161	1.0109	218	1.0111
193	193	1.0270	296	1.0265	192	1.0271	185	1.0197	201	1.0181	146	1.0168	224	1.0153	209	1.0134	186	1.0111	84	1.0106	312	1.0111
194	157	1.0266	175	1.0263	193	1.0271	313	1.0196	312	1.0172	128	1.0167	128	1.0151	207	1.0133	194	1.0111	97	1.0104	185	1.0105
195	312	1.0256	135	1.0241	158	1.0256	144	1.0190	225	1.0170	210	1.0148	312	1.0136	291	1.0132	210	1.0110	170	1.0103	210	1.0104
196	210	1.0253	136	1.0240	185	1.0253	147	1.0174	210	1.0160	201	1.0142	291	1.0129	210	1.0122	207	1.0110	186	1.0103	84	1.0101
197	100	1.0246	158	1.0238	224	1.0246	149	1.0173	310	1.0155	185	1.0142	161	1.0126	85	1.0120	209	1.0110	194	1.0103	207	1.0100
198	296	1.0243	190	1.0237	175	1.0243	158	1.0165	171	1.0151	218	1.0135	308	1.0119	256	1.0102	256	1.0105	256	1.0102	209	1.0100
199	239	1.0236	191	1.0237	149	1.0237	210	1.0160	224	1.0144	178	1.0132	310	1.0117	148	1.0097	178	1.0102	207	1.0100	170	1.0097
200	175	1.0233	119	1.0231	61	1.0234	175	1.0159	168	1.0137	291	1.0130	210	1.0111	73	1.0096	310	1.0102	209	1.0100	256	1.0096
201	224	1.0223	192	1.0224	73	1.0226	208	1.0159	169	1.0137	168	1.0125	256	1.0106	161	1.0095	85	1.0102	85	1.0099	85	1.0092
202	127	1.0204	193	1.0224	217	1.0202	87	1.0157	178	1.0135	169	1.0125	170	1.0105	185	1.0095	122	1.0099	210	1.0098	222	1.0086
203	83	1.0189	124	1.0222	210	1.0201	171	1.0153	148	1.0131	308	1.0114	218	1.0105	218	1.0094	170	1.0091	158	1.0078	223	1.0086
204	292	1.0182	185	1.0217	218	1.0197	159	1.0151	161	1.0127	148	1.0114	148	1.0104	170	1.0087	185	1.0088	218	1.0075	310	1.0082
205	256	1.0181	210	1.0201	100	1.0193	100	1.0140	185	1.0127	159	1.0105	178	1.0103	262	1.0074	148	1.0086	73	1.0068	158	1.0077
206	124	1.0159	208	1.0195	144	1.0190	291	1.0137	158	1.0122	161	1.0104	225	1.0092	178	1.0065	218	1.0079	148	1.0056	217	1.0061
207	217	1.0156	239	1.0194	208	1.0184	224	1.0132	291	1.0121	256	1.0104	73	1.0085	312	1.0059	158	1.0054	222	1.0050	189	1.0054
208	148	1.0147	65	1.0192	296	1.0166	239	1.0130	218	1.0115	73	1.0100	185	1.0080	189	1.0051	189	1.0048	223	1.0050	153	1.0046
209	147	1.0145	66	1.0192	310	1.0152	148	1.0125	256	1.0104	158	1.0097	86	1.0078	149	1.0035	311	1.0047	310	1.0043	149	1.0042
210	140	1.0143	299	1.0190	124	1.0143	73	1.0118	170	1.0100	262	1.0093	311	1.0064	310	1.0033	149	1.0042	122	1.0042	311	1.0028
211	188	1.0131	100	1.0181	239	1.0139	312	1.0116	87	1.0100	170	1.0092	83	1.0056	309	1.0031	159	1.0038	189	1.0041	148	1.0027
212	158	1.0131	49	1.0170	171	1.0124	256	1.0104	149	1.0099	149	1.0085	158	1.0053	290	1.0031	222	1.0031	149	1.0036	309	1.0024

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$
213	185	1.0127	132	1.0160	148	1.0121	161	1.0102	83	1.0098	310	1.0082	189	1.0048	83	1.0025	223	1.0031	309	1.0028	108	1.0010
214	200	1.0125	144	1.0151	83	1.0107	103	1.0102	239	1.0097	83	1.0079	149	1.0045	222	1.0025	309	1.0024	262	1.0024	216	1.0009
215	310	1.0121	83	1.0129	291	1.0102	262	1.0100	208	1.0097	208	1.0067	222	1.0038	223	1.0025	296	1.0023	290	1.0009	110	1.0006
216	171	1.0116	148	1.0127	119	1.0101	83	1.0099	146	1.0092	239	1.0061	223	1.0038	108	1.0023	108	1.0022	108	1.0008	290	1.0003
217	262	1.0096	171	1.0119	65	1.0100	217	1.0096	127	1.0090	296	1.0049	208	1.0033	158	1.0018	188	1.0011	311	1.0004	183	1.0003
218	144	1.0091	310	1.0118	66	1.0100	218	1.0096	308	1.0082	222	1.0044	108	1.0033	239	1.0014	290	1.0011	83	1.0001	28	1.0000
219	311	1.0089	200	1.0116	161	1.0094	310	1.0086	311	1.0078	223	1.0044	239	1.0031	50	1.0010	83	1.0009	216	1.0001	65	1.0000
220	263	1.0088	300	1.0109	110	1.0092	170	1.0084	296	1.0066	189	1.0043	309	1.0030	188	1.0008	28	1.0000	110	1.0001	227	1.0000
221	228	1.0086	61	1.0105	297	1.0091	124	1.0079	123	1.0050	311	1.0042	290	1.0024	311	1.0005	65	1.0000	28	1.0000	228	1.0000
222	229	1.0086	262	1.0099	298	1.0091	299	1.0075	189	1.0050	127	1.0034	296	1.0020	28	1.0000	227	1.0000	65	1.0000	102	1.0000
223	142	1.0086	256	1.0095	311	1.0090	296	1.0067	188	1.0048	290	1.0034	188	1.0019	65	1.0000	228	1.0000	227	1.0000	83	1.0000
224	227	1.0083	170	1.0090	256	1.0089	189	1.0044	290	1.0046	309	1.0026	124	1.0011	227	1.0000	102	1.0000	228	1.0000	229	1.0000
225	161	1.0075	291	1.0078	188	1.0076	188	1.0042	100	1.0038	188	1.0025	201	1.0010	228	1.0000	229	1.0000	102	1.0000	142	1.0000
226	170	1.0065	188	1.0068	170	1.0073	146	1.0035	262	1.0029	108	1.0017	262	1.0006	102	1.0000	142	1.0000	229	1.0000	66	1.0000
227	300	1.0048	297	1.0067	168	1.0068	308	1.0031	222	1.0026	87	1.0004	28	1.0000	229	1.0000	66	1.0000	142	1.0000	73	1.0000
228	290	1.0045	298	1.0067	169	1.0068	290	1.0030	223	1.0026	228	1.0002	65	1.0000	142	1.0000	73	1.0000	66	1.0000	292	1.0000
229	238	1.0026	311	1.0044	262	1.0065	311	1.0027	175	1.0021	110	1.0002	227	1.0000	66	1.0000	292	1.0000	292	1.0000	206	1.0000
230	108	1.0014	161	1.0041	290	1.0036	227	1.0020	309	1.0021	227	1.0001	228	1.0000	292	1.0000	206	1.0000	206	1.0000	230	1.0000
231	206	1.0004	140	1.0041	189	1.0031	228	1.0020	108	1.0020	229	1.0001	102	1.0000	206	1.0000	230	1.0000	230	1.0000	231	1.0000
232	179	1.0003	263	1.0040	222	1.0030	229	1.0020	103	1.0006	28	1.0000	229	1.0000	230	1.0000	231	1.0000	231	1.0000	232	1.0000
233	28	1.0000	189	1.0036	223	1.0030	168	1.0011	124	1.0004	65	1.0000	142	1.0000	231	1.0000	232	1.0000	232	1.0000	233	1.0000
234	146	1.0000	108	1.0033	299	1.0021	169	1.0011	228	1.0003	102	1.0000	66	1.0000	232	1.0000	233	1.0000	233	1.0000	49	1.0000
235	230	1.0000	290	1.0024	108	1.0018	119	1.0008	227	1.0002	142	1.0000	292	1.0000	233	1.0000	49	1.0000	49	1.0000	234	1.0000
236	231	1.0000	179	1.0022	308	1.0017	300	1.0008	229	1.0002	66	1.0000	206	1.0000	49	1.0000	234	1.0000	234	1.0000	235	1.0000
237	232	1.0000	222	1.0021	229	1.0013	65	1.0006	235	1.0001	292	1.0000	230	1.0000	234	1.0000	235	1.0000	235	1.0000	236	1.0000
238	233	1.0000	223	1.0021	300	1.0013	66	1.0006	244	1.0001	206	1.0000	231	1.0000	235	1.0000	236	1.0000	236	1.0000	237	1.0000
239	49	1.0000	238	1.0015	227	1.0012	108	1.0005	110	1.0001	230	1.0000	232	1.0000	236	1.0000	237	1.0000	237	1.0000	238	1.0000
240	234	1.0000	28	1.0000	228	1.0012	235	1.0001	28	1.0000	231	1.0000	233	1.0000	237	1.0000	238	1.0000	238	1.0000	239	1.0000
241	236	1.0000	102	1.0000	200	1.0007	244	1.0001	65	1.0000	232	1.0000	234	1.0000	238	1.0000	239	1.0000	239	1.0000	240	1.0000
242	237	1.0000	142	1.0000	146	1.0003	28	1.0000	102	1.0000	233	1.0000	235	1.0000	240	1.0000	240	1.0000	240	1.0000	241	1.0000
243	240	1.0000	206	1.0000	263	1.0002	102	1.0000	142	1.0000	234	1.0000	236	1.0000	241	1.0000	241	1.0000	241	1.0000	119	1.0000
244	241	1.0000	146	1.0000	28	1.0000	142	1.0000	66	1.0000	235	1.0000	237	1.0000	119	1.0000	119	1.0000	119	1.0000	242	1.0000
245	242	1.0000	230	1.0000	102	1.0000	292	1.0000	292	1.0000	236	1.0000	238	1.0000	242	1.0000	242	1.0000	242	1.0000	179	1.0000
246	243	1.0000	231	1.0000	142	1.0000	206	1.0000	206	1.0000	237	1.0000	240	1.0000	179	1.0000	179	1.0000	179	1.0000	50	1.0000
247	245	1.0000	232	1.0000	292	1.0000	230	1.0000	230	1.0000	238	1.0000	241	1.0000	103	1.0000	50	1.0000	50	1.0000	103	1.0000
248	247	1.0000	233	1.0000	206	1.0000	231	1.0000	231	1.0000	240	1.0000	119	1.0000	243	1.0000	103	1.0000	103	1.0000	243	1.0000
249	248	1.0000	234	1.0000	230	1.0000	232	1.0000	232	1.0000	241	1.0000	242	1.0000	244	1.0000	243	1.0000	243	1.0000	244	1.0000
250	249	1.0000	236	1.0000	231	1.0000	233	1.0000	233	1.0000	119	1.0000	179	1.0000	245	1.0000	244	1.0000	244	1.0000	245	1.0000
251	250	1.0000	237	1.0000	232	1.0000	234	1.0000	234	1.0000	242	1.0000	103	1.0000	247	1.0000	245	1.0000	245	1.0000	247	1.0000
252	251	1.0000	240	1.0000	233	1.0000	236	1.0000	236	1.0000	179	1.0000	243	1.0000	248	1.0000	247	1.0000	247	1.0000	248	1.0000
253	252	1.0000	241	1.0000	234	1.0000	237	1.0000	237	1.0000	103	1.0000	244	1.0000	249	1.0000	248	1.0000	248	1.0000	249	1.0000
254	253	1.0000	242	1.0000	236	1.0000	238	1.0000	238	1.0000	243	1.0000	245	1.0000	250	1.0000	249	1.0000	249	1.0000	250	1.0000
255	254	1.0000	243	1.0000	237	1.0000	240	1.0000	240	1.0000	244	1.0000	247	1.0000	251	1.0000	250	1.0000	250	1.0000	251	1.0000
256	255	1.0000	245	1.0000	238	1.0000	241	1.0000	241	1.0000	245	1.0000	248	1.0000	252	1.0000	251	1.0000	251	1.0000	252	1.0000

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$	k	$I(\lambda;k)$
257	257	1.0000	247	1.0000	240	1.0000	242	1.0000	119	1.0000	247	1.0000	249	1.0000	253	1.0000	252	1.0000	252	1.0000	253	1.0000
258	258	1.0000	248	1.0000	241	1.0000	179	1.0000	242	1.0000	248	1.0000	250	1.0000	254	1.0000	253	1.0000	253	1.0000	254	1.0000
259	259	1.0000	249	1.0000	242	1.0000	243	1.0000	179	1.0000	249	1.0000	251	1.0000	255	1.0000	254	1.0000	254	1.0000	255	1.0000
260	260	1.0000	250	1.0000	179	1.0000	245	1.0000	243	1.0000	250	1.0000	252	1.0000	257	1.0000	255	1.0000	255	1.0000	257	1.0000
261	261	1.0000	251	1.0000	243	1.0000	247	1.0000	245	1.0000	251	1.0000	253	1.0000	258	1.0000	257	1.0000	257	1.0000	258	1.0000
262	264	1.0000	252	1.0000	245	1.0000	248	1.0000	247	1.0000	252	1.0000	254	1.0000	259	1.0000	258	1.0000	258	1.0000	259	1.0000
263	265	1.0000	253	1.0000	247	1.0000	249	1.0000	248	1.0000	253	1.0000	255	1.0000	260	1.0000	259	1.0000	259	1.0000	260	1.0000
264	266	1.0000	254	1.0000	248	1.0000	250	1.0000	249	1.0000	254	1.0000	257	1.0000	261	1.0000	260	1.0000	260	1.0000	261	1.0000
265	267	1.0000	255	1.0000	249	1.0000	251	1.0000	250	1.0000	255	1.0000	258	1.0000	216	1.0000	261	1.0000	261	1.0000	262	1.0000
266	268	1.0000	257	1.0000	250	1.0000	252	1.0000	251	1.0000	257	1.0000	259	1.0000	263	1.0000	216	1.0000	263	1.0000	263	1.0000
267	269	1.0000	258	1.0000	251	1.0000	253	1.0000	252	1.0000	258	1.0000	260	1.0000	264	1.0000	262	1.0000	264	1.0000	264	1.0000
268	270	1.0000	259	1.0000	252	1.0000	254	1.0000	253	1.0000	259	1.0000	261	1.0000	265	1.0000	263	1.0000	265	1.0000	265	1.0000
269	271	1.0000	260	1.0000	253	1.0000	255	1.0000	254	1.0000	260	1.0000	263	1.0000	266	1.0000	264	1.0000	266	1.0000	266	1.0000
270	272	1.0000	261	1.0000	254	1.0000	257	1.0000	255	1.0000	261	1.0000	264	1.0000	267	1.0000	265	1.0000	267	1.0000	267	1.0000
271	273	1.0000	264	1.0000	255	1.0000	258	1.0000	257	1.0000	263	1.0000	265	1.0000	268	1.0000	266	1.0000	268	1.0000	268	1.0000
272	274	1.0000	265	1.0000	257	1.0000	259	1.0000	258	1.0000	264	1.0000	266	1.0000	269	1.0000	267	1.0000	269	1.0000	269	1.0000
273	275	1.0000	266	1.0000	258	1.0000	260	1.0000	259	1.0000	265	1.0000	267	1.0000	270	1.0000	268	1.0000	270	1.0000	270	1.0000
274	276	1.0000	267	1.0000	259	1.0000	261	1.0000	260	1.0000	266	1.0000	268	1.0000	271	1.0000	269	1.0000	271	1.0000	271	1.0000
275	277	1.0000	268	1.0000	260	1.0000	263	1.0000	261	1.0000	267	1.0000	269	1.0000	272	1.0000	270	1.0000	272	1.0000	272	1.0000
276	278	1.0000	269	1.0000	261	1.0000	264	1.0000	263	1.0000	268	1.0000	270	1.0000	273	1.0000	271	1.0000	273	1.0000	273	1.0000
277	279	1.0000	270	1.0000	264	1.0000	265	1.0000	264	1.0000	269	1.0000	271	1.0000	274	1.0000	272	1.0000	274	1.0000	274	1.0000
278	280	1.0000	271	1.0000	265	1.0000	266	1.0000	265	1.0000	270	1.0000	272	1.0000	275	1.0000	273	1.0000	275	1.0000	275	1.0000
279	281	1.0000	272	1.0000	266	1.0000	267	1.0000	266	1.0000	271	1.0000	273	1.0000	276	1.0000	274	1.0000	276	1.0000	276	1.0000
280	282	1.0000	273	1.0000	267	1.0000	268	1.0000	267	1.0000	272	1.0000	274	1.0000	277	1.0000	275	1.0000	277	1.0000	277	1.0000
281	283	1.0000	274	1.0000	268	1.0000	269	1.0000	268	1.0000	273	1.0000	275	1.0000	278	1.0000	276	1.0000	278	1.0000	278	1.0000
282	284	1.0000	275	1.0000	269	1.0000	270	1.0000	269	1.0000	274	1.0000	276	1.0000	279	1.0000	277	1.0000	279	1.0000	279	1.0000
283	285	1.0000	276	1.0000	270	1.0000	271	1.0000	270	1.0000	275	1.0000	277	1.0000	280	1.0000	278	1.0000	280	1.0000	280	1.0000
284	286	1.0000	277	1.0000	271	1.0000	272	1.0000	271	1.0000	276	1.0000	278	1.0000	281	1.0000	279	1.0000	281	1.0000	281	1.0000
285	287	1.0000	278	1.0000	272	1.0000	273	1.0000	272	1.0000	277	1.0000	279	1.0000	282	1.0000	280	1.0000	282	1.0000	282	1.0000
286	288	1.0000	279	1.0000	273	1.0000	274	1.0000	273	1.0000	278	1.0000	280	1.0000	283	1.0000	281	1.0000	283	1.0000	283	1.0000
287	135	0.9992	280	1.0000	274	1.0000	275	1.0000	274	1.0000	279	1.0000	281	1.0000	284	1.0000	282	1.0000	284	1.0000	284	1.0000
288	189	0.9992	281	1.0000	275	1.0000	276	1.0000	275	1.0000	280	1.0000	282	1.0000	140	1.0000	283	1.0000	140	1.0000	140	1.0000
289	244	0.9978	282	1.0000	276	1.0000	277	1.0000	276	1.0000	281	1.0000	283	1.0000	183	1.0000	284	1.0000	183	1.0000	285	1.0000
290	168	0.9977	283	1.0000	277	1.0000	278	1.0000	277	1.0000	282	1.0000	284	1.0000	285	1.0000	140	1.0000	285	1.0000	286	1.0000
291	169	0.9977	284	1.0000	278	1.0000	279	1.0000	278	1.0000	283	1.0000	140	1.0000	286	1.0000	183	1.0000	286	1.0000	87	1.0000
292	235	0.9971	285	1.0000	279	1.0000	280	1.0000	279	1.0000	284	1.0000	183	1.0000	87	1.0000	285	1.0000	87	1.0000	287	1.0000
293	223	0.9971	286	1.0000	280	1.0000	281	1.0000	280	1.0000	140	1.0000	285	1.0000	110	1.0000	286	1.0000	287	1.0000	288	1.0000
294	222	0.9970	287	1.0000	281	1.0000	282	1.0000	281	1.0000	183	1.0000	286	1.0000	287	1.0000	87	1.0000	288	1.0000	201	0.9990
295	291	0.9965	288	1.0000	282	1.0000	283	1.0000	282	1.0000	285	1.0000	87	1.0000	288	1.0000	110	1.0000	127	0.9994	188	0.9988
296	309	0.9942	63	0.9970	283	1.0000	284	1.0000	283	1.0000	286	1.0000	110	1.0000	124	0.9982	287	1.0000	188	0.9987	97	0.9987
297	297	0.9938	229	0.9968	284	1.0000	140	1.0000	284	1.0000	287	1.0000	287	1.0000	208	0.9972	288	1.0000	200	0.9975	127	0.9985
298	298	0.9938	227	0.9965	140	1.0000	285	1.0000	140	1.0000	288	1.0000	288	1.0000	299	0.9969	127	0.9999	208	0.9959	200	0.9969
299	299	0.9919	228	0.9964	285	1.0000	286	1.0000	183	1.0000	175	0.9996	127	0.9977	300	0.9947	124	0.9978	124	0.9958	314	0.9962
300	308	0.9892	244	0.9963	286	1.0000	287	1.0000	285	1.0000	124	0.9978	216	0.9965	296	0.9943	225	0.9972	296	0.9951	296	0.9951

Order	$n = 10$		$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$n = 20$	
	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$	k	$I(\lambda; k)$
301	136	0.9887	235	0.9961	287	1.0000	288	1.0000	286	1.0000	300	0.9976	314	0.9953	127	0.9937	314	0.9957	300	0.9947	300	0.9949
302	314	0.9860	308	0.9948	288	1.0000	309	0.9994	287	1.0000	200	0.9964	295	0.9950	314	0.9912	300	0.9956	201	0.9939	124	0.9932
303	302	0.9856	309	0.9946	244	0.9997	222	0.9990	288	1.0000	299	0.9943	300	0.9949	159	0.9892	208	0.9950	314	0.9927	208	0.9906
304	47	0.9840	201	0.9917	235	0.9996	223	0.9990	147	0.9993	216	0.9941	200	0.9913	201	0.9885	200	0.9924	175	0.9924	175	0.9870
305	304	0.9827	314	0.9906	309	0.9982	110	0.9964	314	0.9981	314	0.9930	159	0.9894	200	0.9882	175	0.9906	299	0.9868	225	0.9857
306	201	0.9816	168	0.9905	314	0.9933	200	0.9942	300	0.9977	217	0.9925	175	0.9892	175	0.9880	201	0.9861	147	0.9833	299	0.9821
307	63	0.9804	169	0.9905	132	0.9853	183	0.9916	200	0.9928	147	0.9912	299	0.9877	217	0.9807	299	0.9848	159	0.9813	147	0.9805
308	301	0.9801	47	0.9904	47	0.9842	314	0.9902	299	0.9919	100	0.9818	147	0.9810	147	0.9759	147	0.9811	304	0.9765	304	0.9740
309	132	0.9796	304	0.9885	183	0.9815	216	0.9737	216	0.9845	302	0.9742	47	0.9750	302	0.9750	302	0.9727	302	0.9765	302	0.9740
310	183	0.9767	302	0.9847	302	0.9779	47	0.9728	47	0.9758	47	0.9737	302	0.9733	304	0.9749	304	0.9726	217	0.9751	159	0.9638
311	102	0.9758	301	0.9825	304	0.9775	302	0.9728	302	0.9728	304	0.9735	304	0.9731	47	0.9696	47	0.9698	47	0.9671	47	0.9611
312	37	0.9725	37	0.9759	301	0.9690	304	0.9722	304	0.9726	295	0.9580	100	0.9597	100	0.9576	100	0.9539	100	0.9566	100	0.9497
313	61	0.9668	183	0.9743	216	0.9681	301	0.9602	301	0.9583	301	0.9580	301	0.9531	301	0.9536	301	0.9492	301	0.9548	301	0.9488
314	216	0.9412	216	0.9302	37	0.9622	37	0.9539	37	0.9526	37	0.9529	37	0.9487	37	0.9503	37	0.9452	37	0.9510	37	0.9472
315	315	NA	315	NA	315	NA	315	NA	315	NA	315	NA	315	NA	315	NA	315	NA	315	NA	315	NA

NA = not available (due to the value of the invariant *per se* is an indeterminate value, with division by zero).

APÊNDICE D – Gráficos dos invariantes das Tabelas 2 e 3

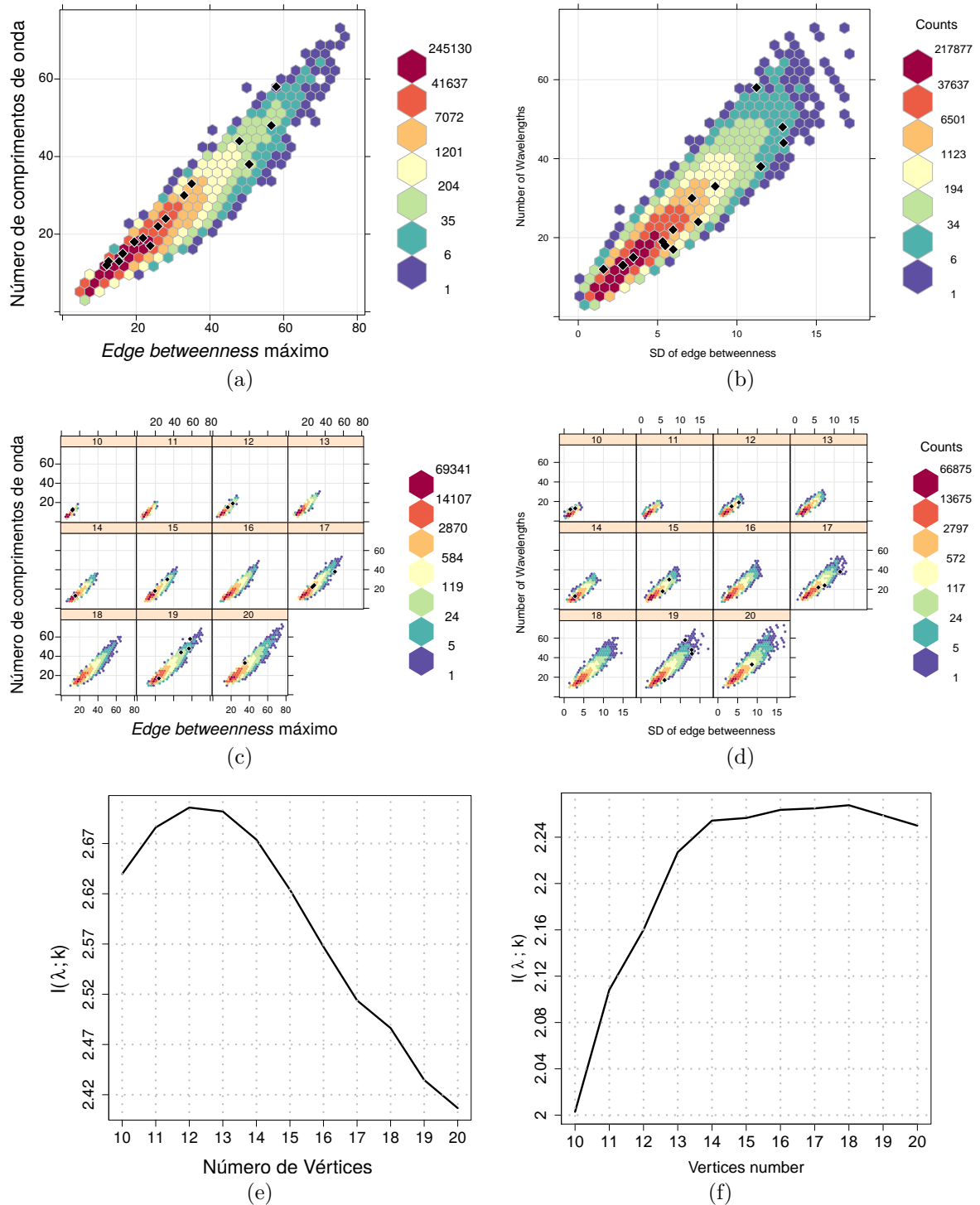


Figura 28 – Maximum of edge betweenness ($k = 1$) and Standard deviation of edge betweenness ($k = 2$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum of edge betweenness and SD of edge betweenness both with λ , as a function of network order.

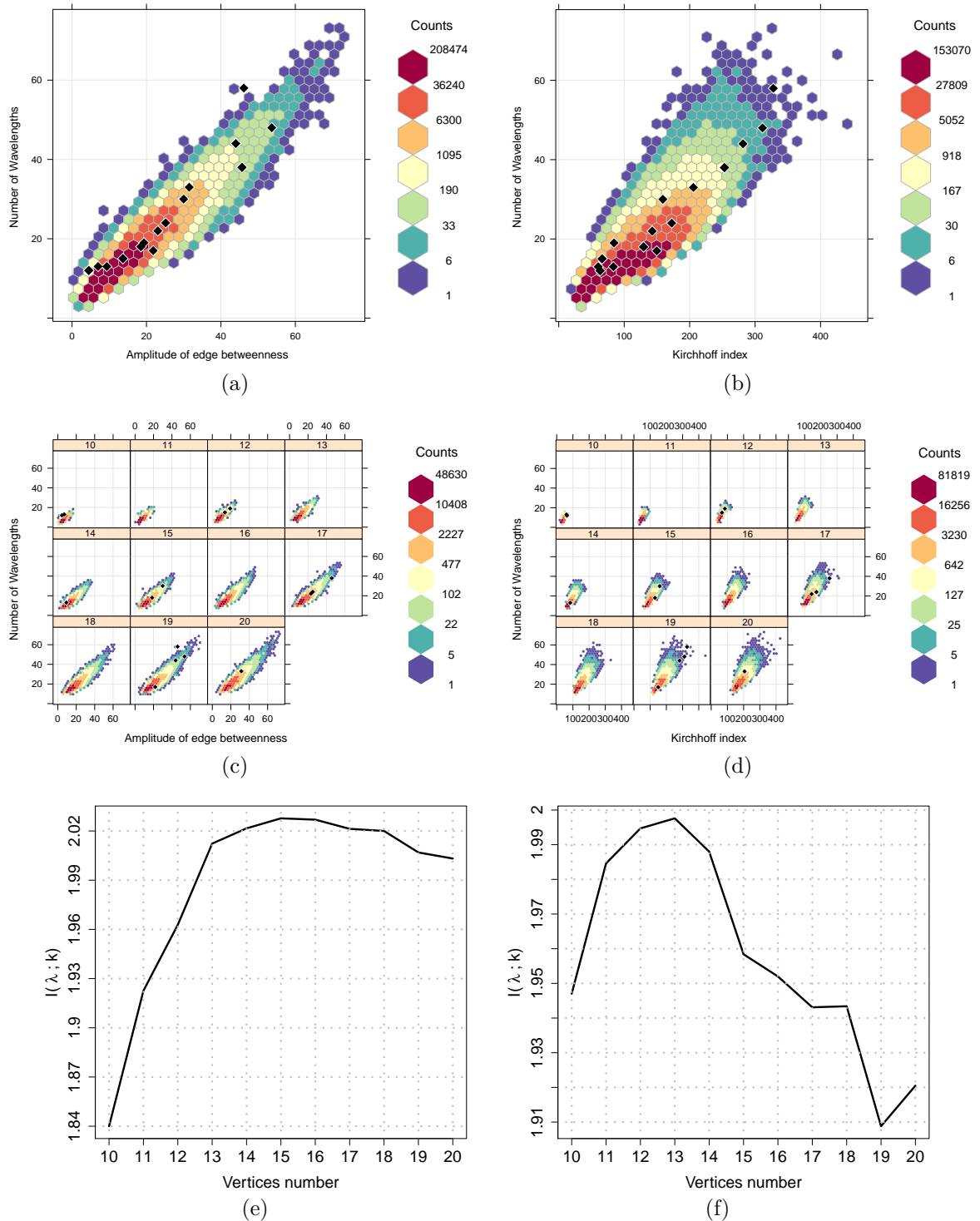


Figura 29 – Amplitude of edge betweenness ($k = 3$) and Kirchhoff index ($k = 4$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Amplitude of edge betweenness and Kirchhoff index both with λ , as a function of network order.

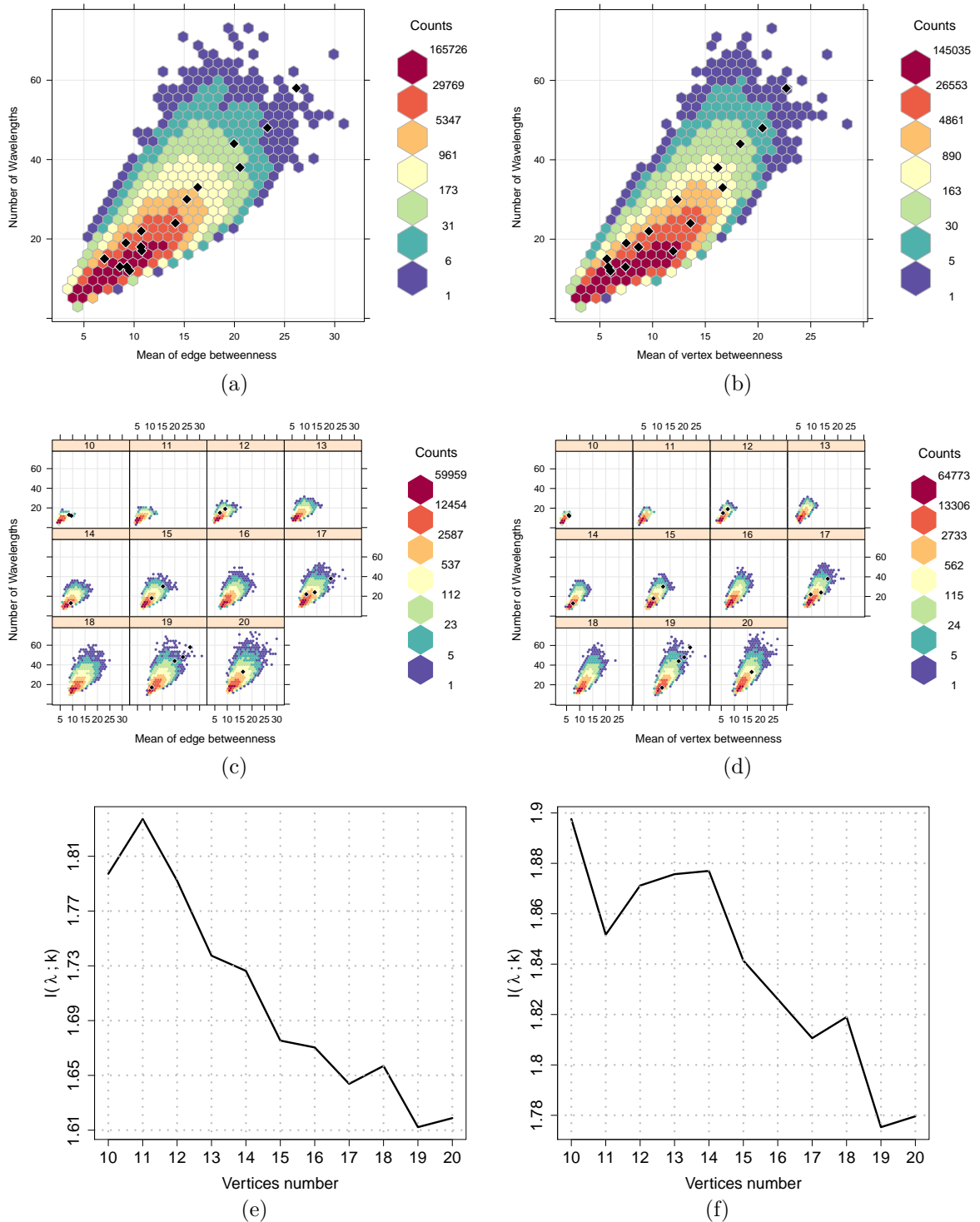


Figura 30 – Mean of edge betweenness ($k = 5$) and Mean of vertex betweenness ($k = 6$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean of edge betweenness and Mean of vertex betweenness both with λ , as a function of network order.

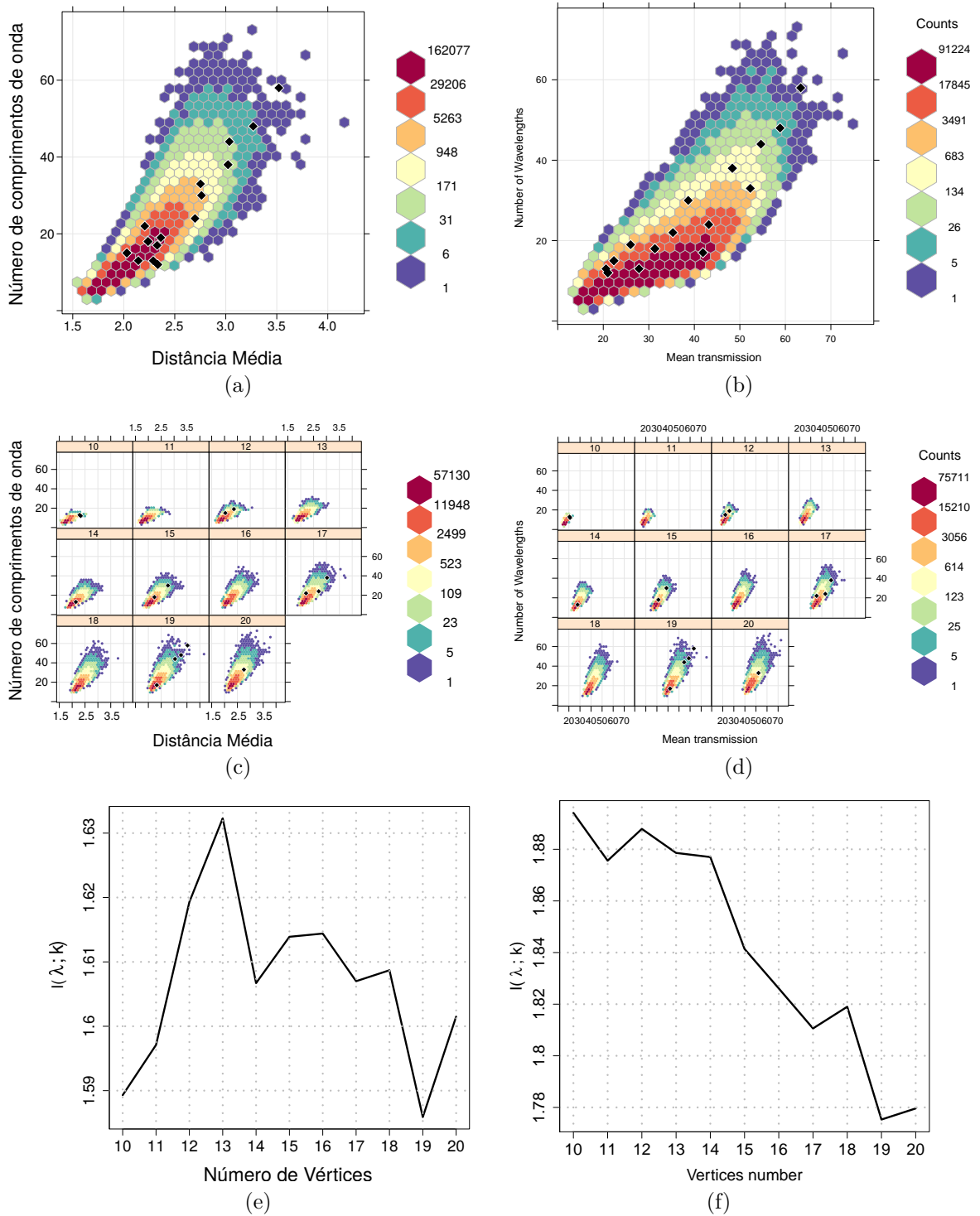


Figura 31 – Mean distance ($k = 7$) and Mean transmission ($k = 8$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean distance and Mean transmission both with λ , as a function of network order.

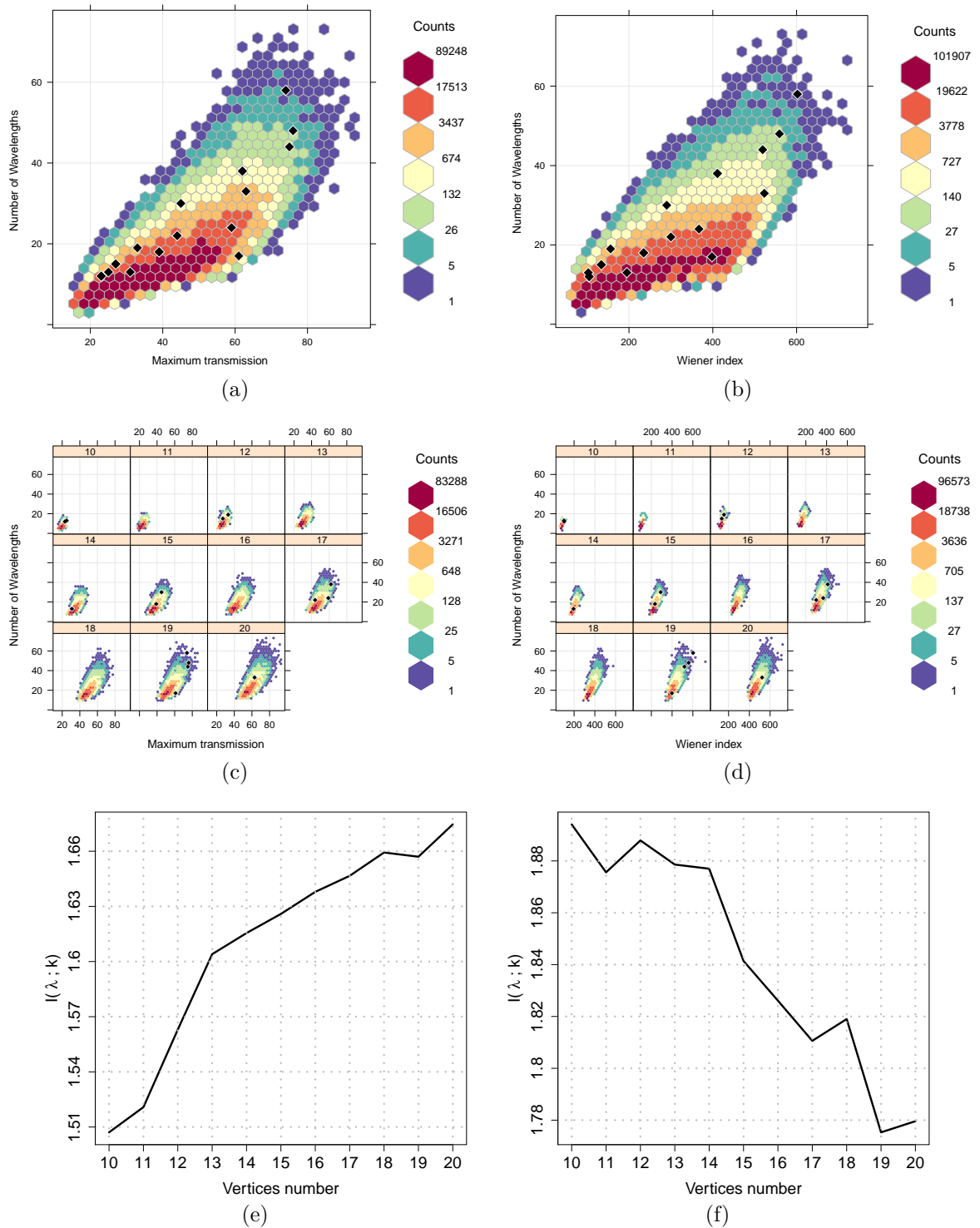


Figura 32 – Maximum transmission ($k = 9$) and Wiener index ($k = 10$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum transmission and Wiener index both with λ , as a function of network order.

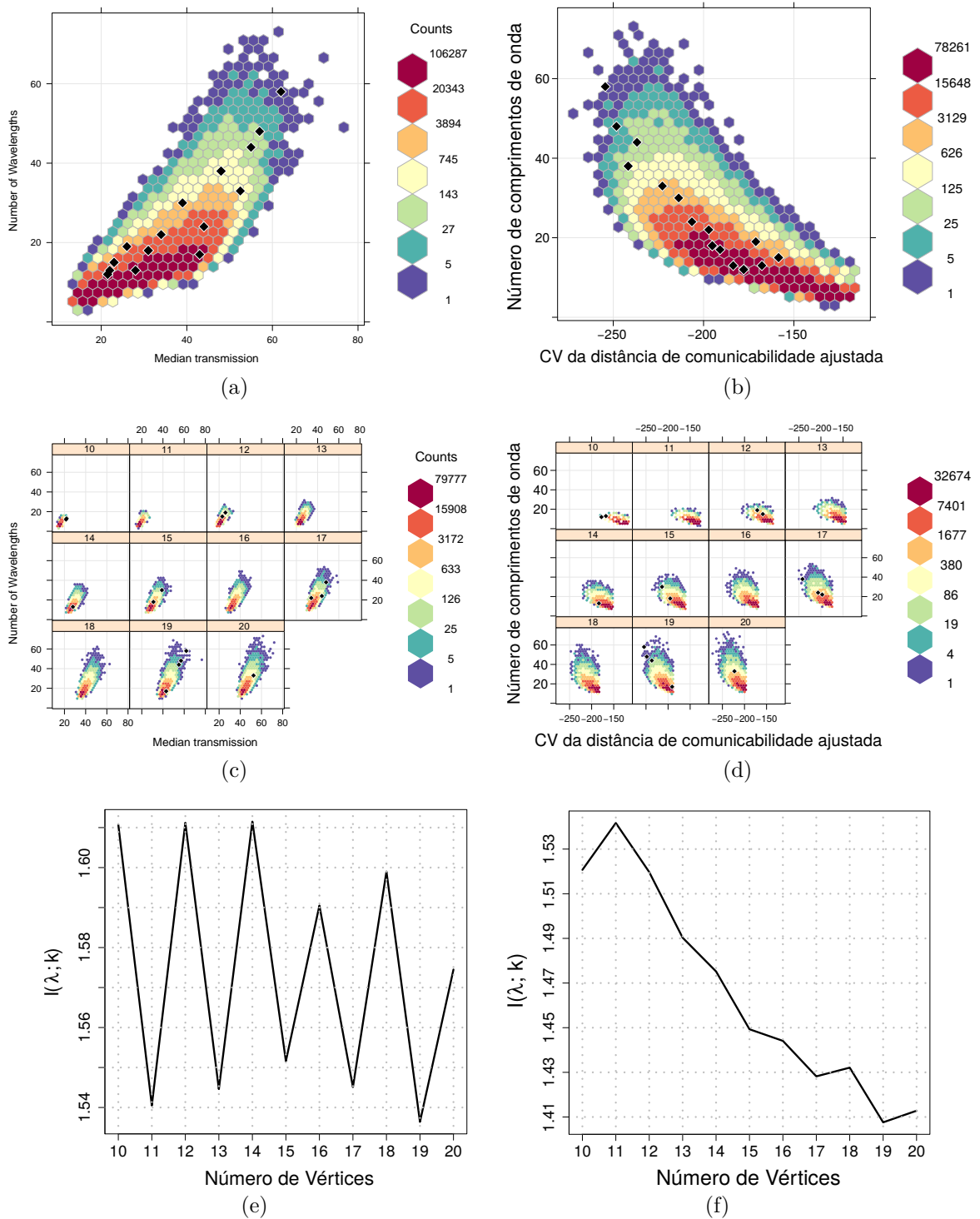


Figura 33 – Median transmission ($k = 11$) and Coefficient of variation of adjusted communicability distance ($k = 12$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Median transmission and Coefficient of variation of adjusted communicability distance both with λ , as a function of network order.

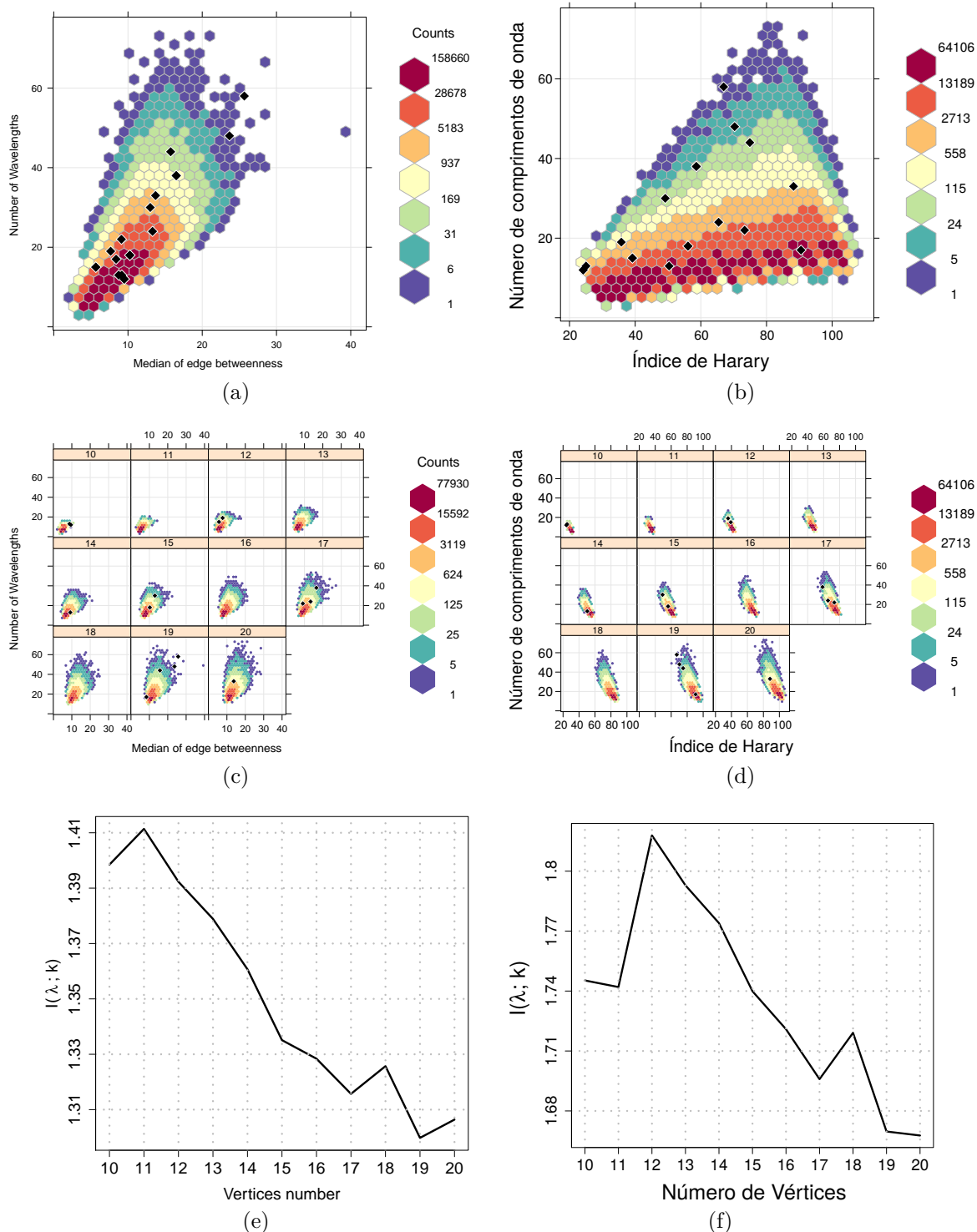


Figura 34 – Median of edge betweenness ($k = 13$) and Harary index ($k = 14$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Median of edge betweenness and Harary index both with λ , as a function of network order.

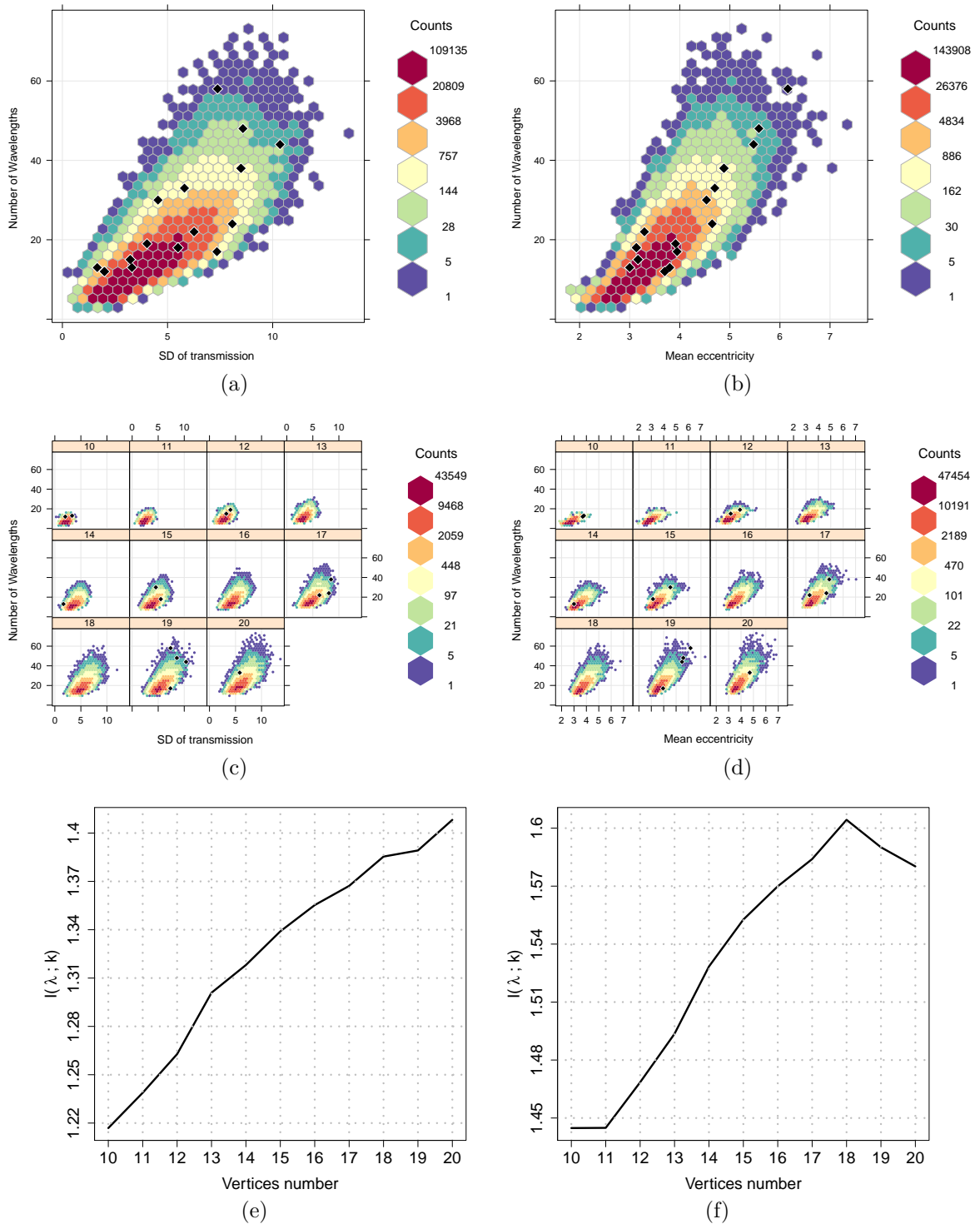


Figura 35 – Standard deviation of transmission ($k = 15$) and Mean eccentricity ($k = 16$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Standard deviation of transmission and Mean eccentricity both with λ , as a function of network order.

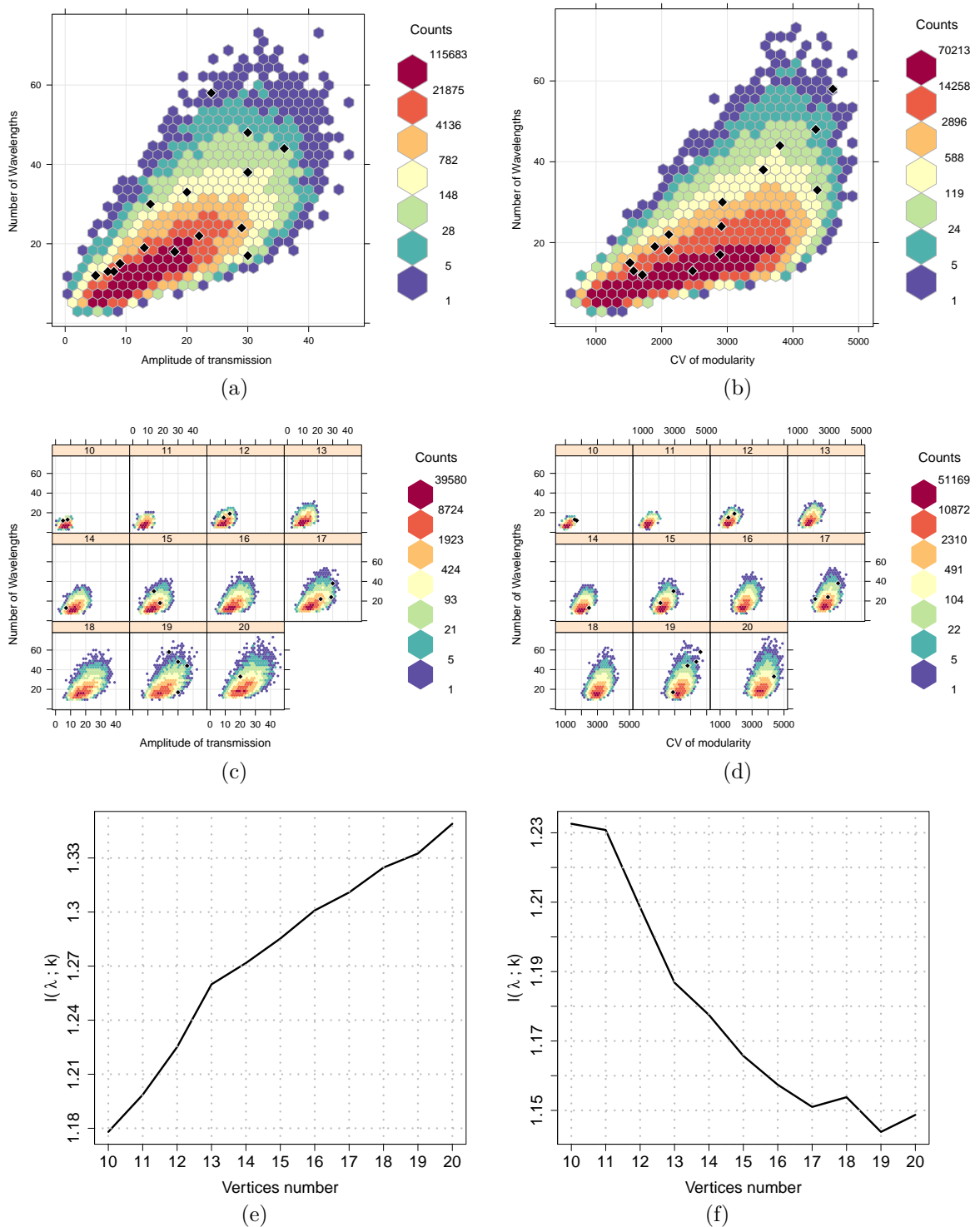


Figura 36 – Amplitude of transmission ($k = 17$) and Coefficient of variation of modularity ($k = 18$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Amplitude of transmission and Coefficient of variation of modularity both with λ , as a function of network order.

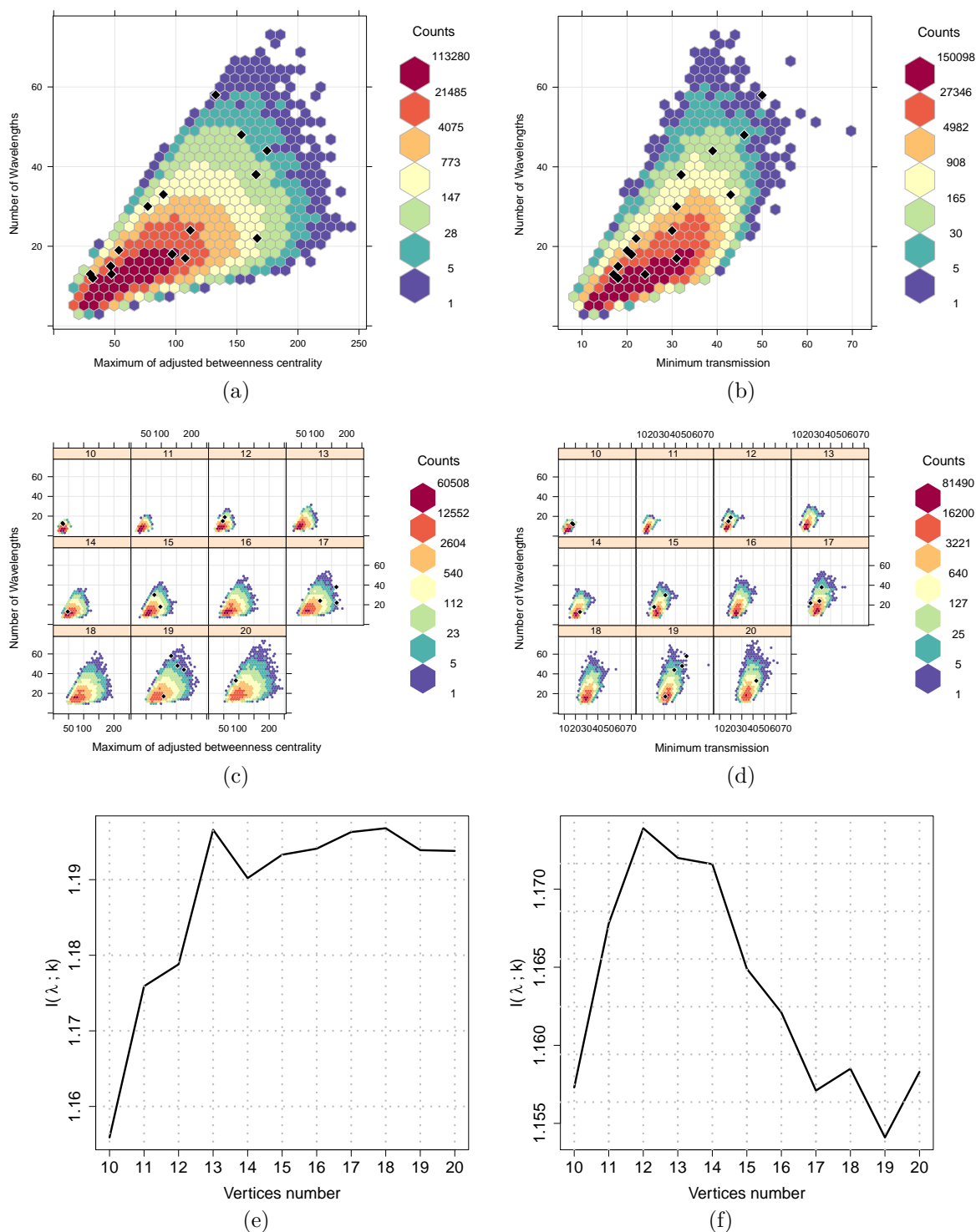


Figura 37 – Maximum of adjusted betweenness centrality ($k = 19$) and Minimum transmission ($k = 20$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum of adjusted betweenness centrality and Minimum transmission both with λ , as a function of network order.

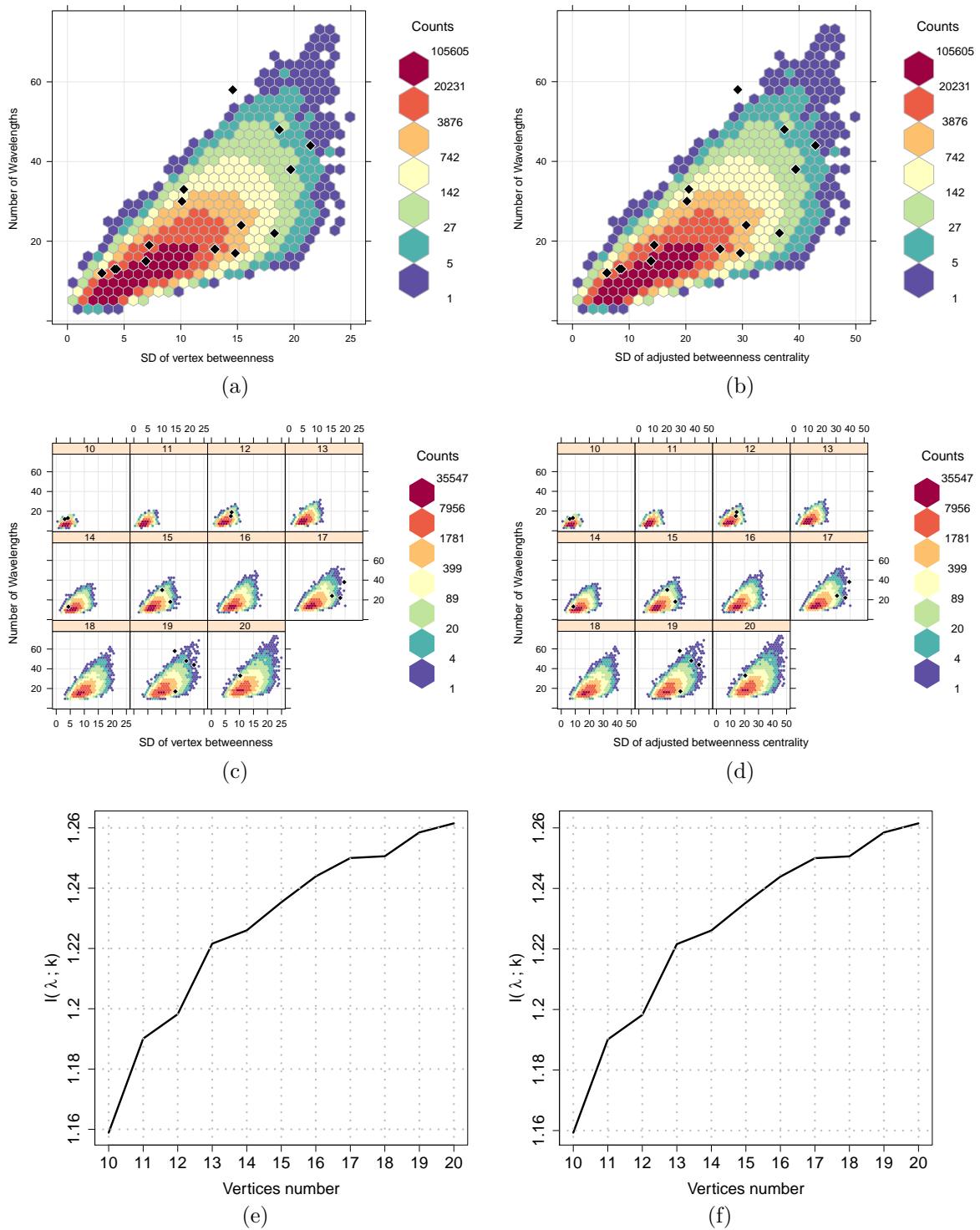


Figura 38 – Standard deviation of vertex betweenness ($k = 21$) and Standard deviation of adjusted betweenness centrality ($k = 22$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Standard deviation of vertex betweenness and Standard deviation of adjusted betweenness centrality both with λ , as a function of network order.

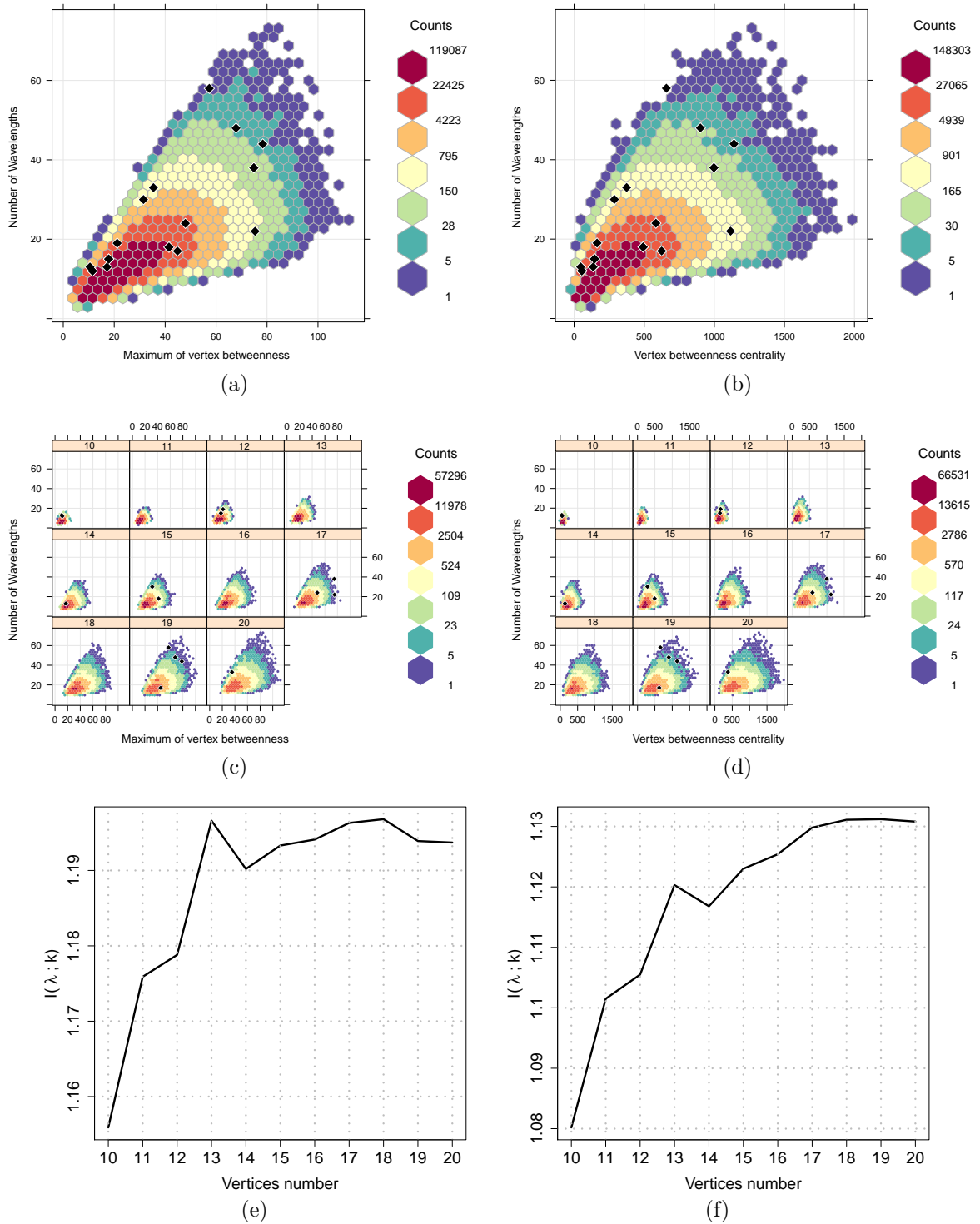


Figura 39 – Maximum of vertex betweenness and Vertex betweenness centrality both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum of vertex betweenness and Vertex betweenness centrality both with λ , as a function of network order.

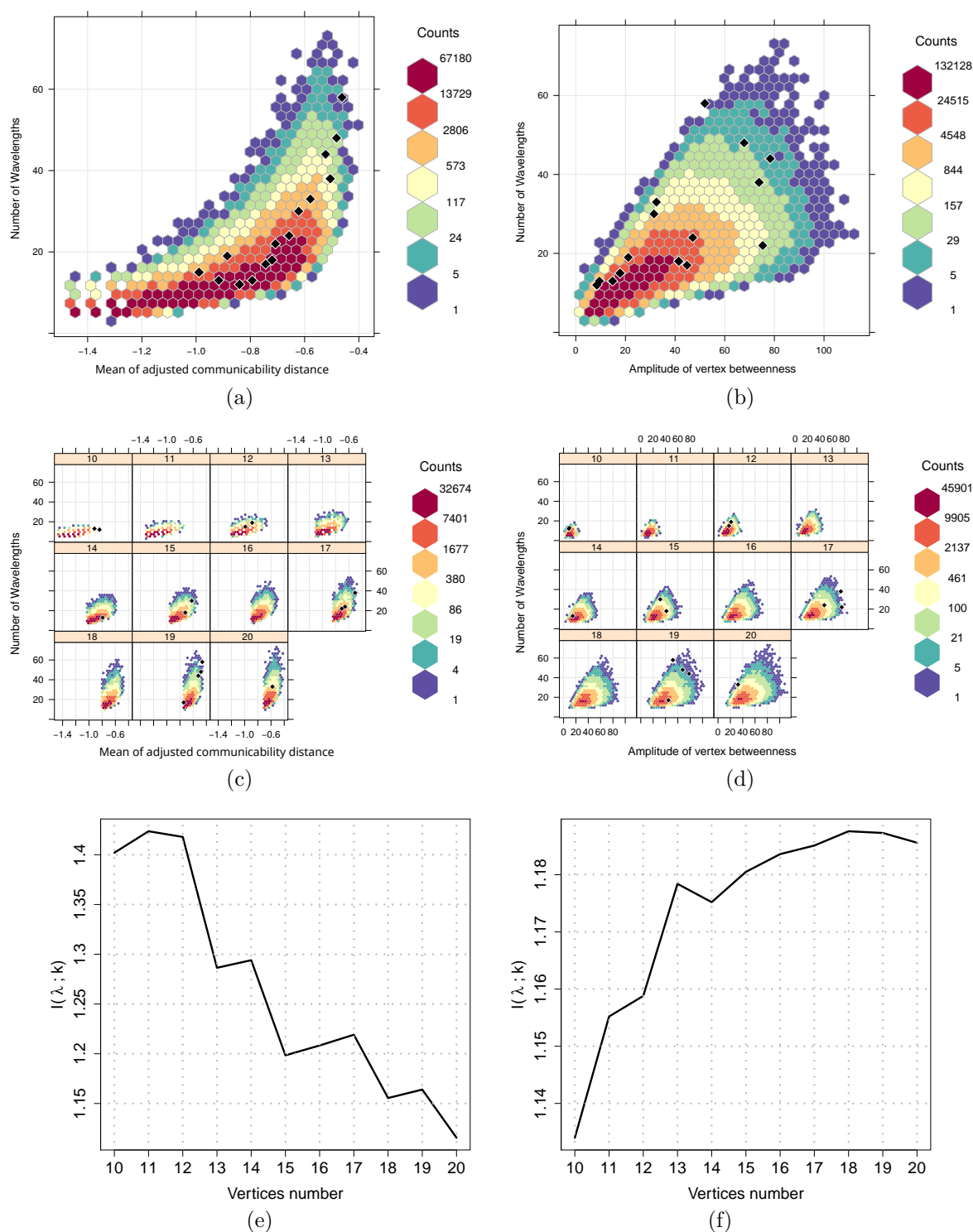


Figura 40 – Mean of adjusted communicability distance ($k = 25$) and Amplitude of vertex betweenness ($k = 26$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean of adjusted communicability distance and Amplitude of vertex betweenness both with λ , as a function of network order.

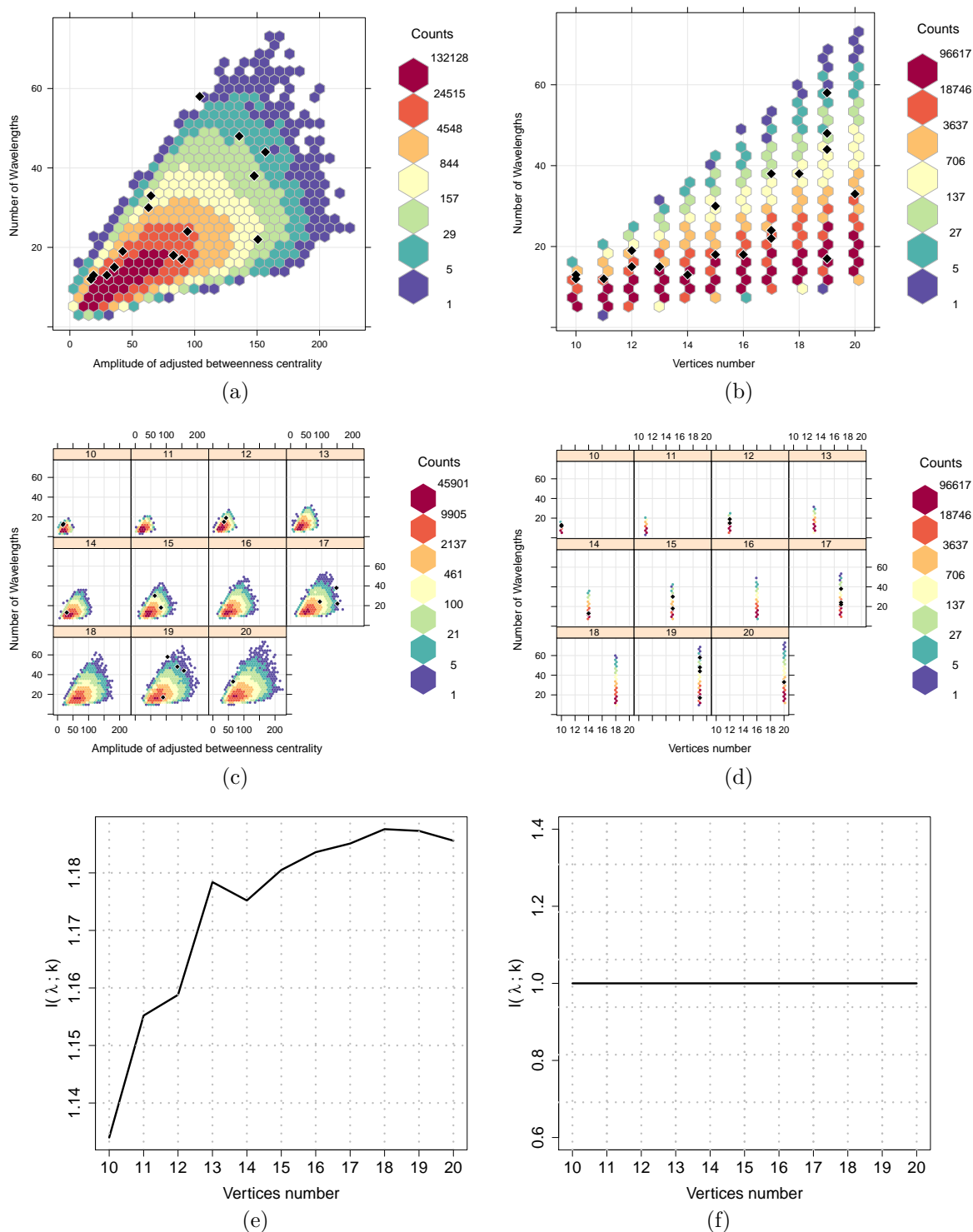


Figura 41 – Amplitude of adjusted betweenness centrality ($k = 27$) and Vertices number ($k = 28$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Amplitude of adjusted betweenness centrality and Vertices number both with λ , as a function of network order.

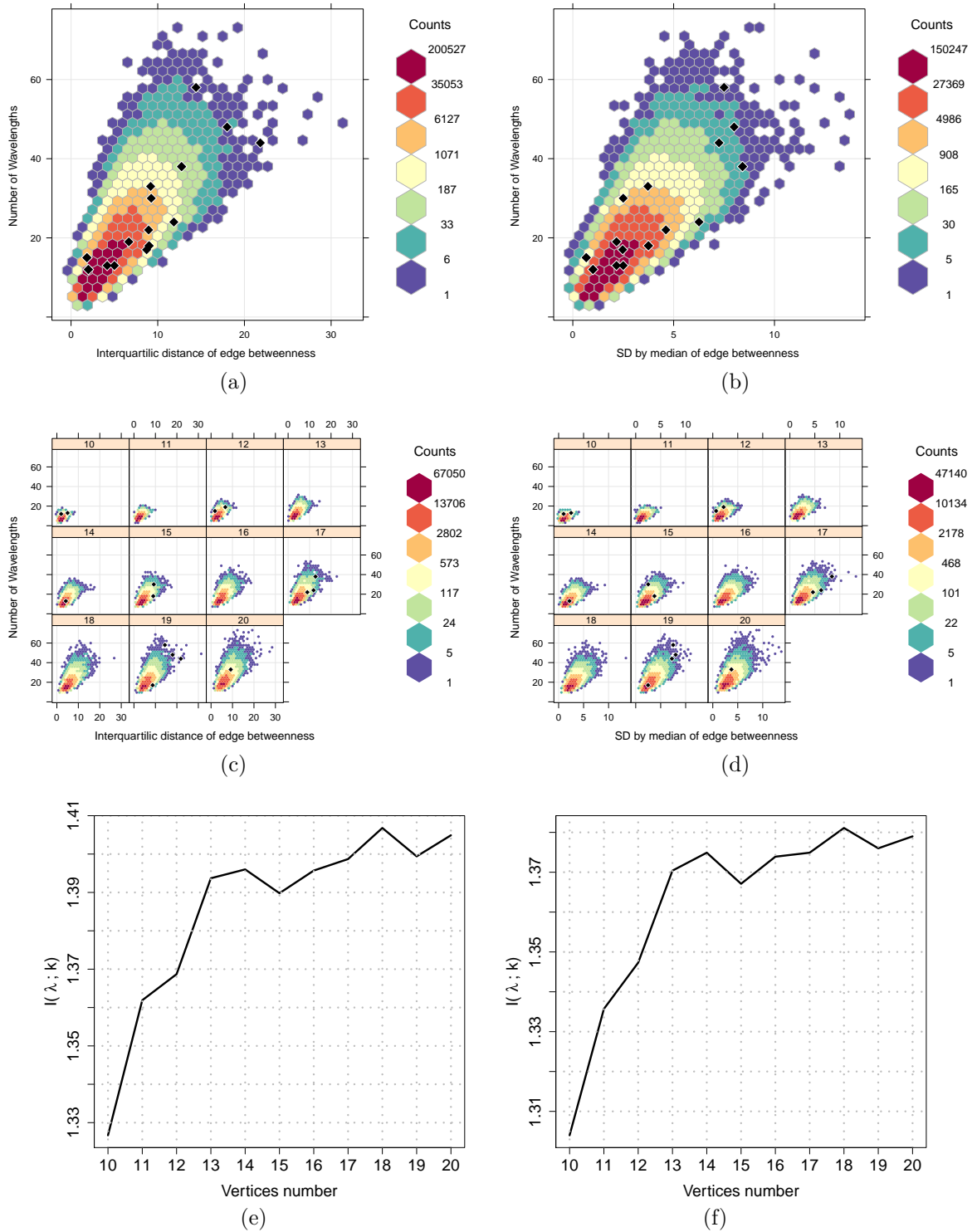


Figura 42 – Interquartilic distance of edge betweenness ($k = 29$) and Standard deviation by median of edge betweenness ($k = 30$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Interquartilic distance of edge betweenness and Standard deviation by median of edge betweenness both with λ , as a function of network order.

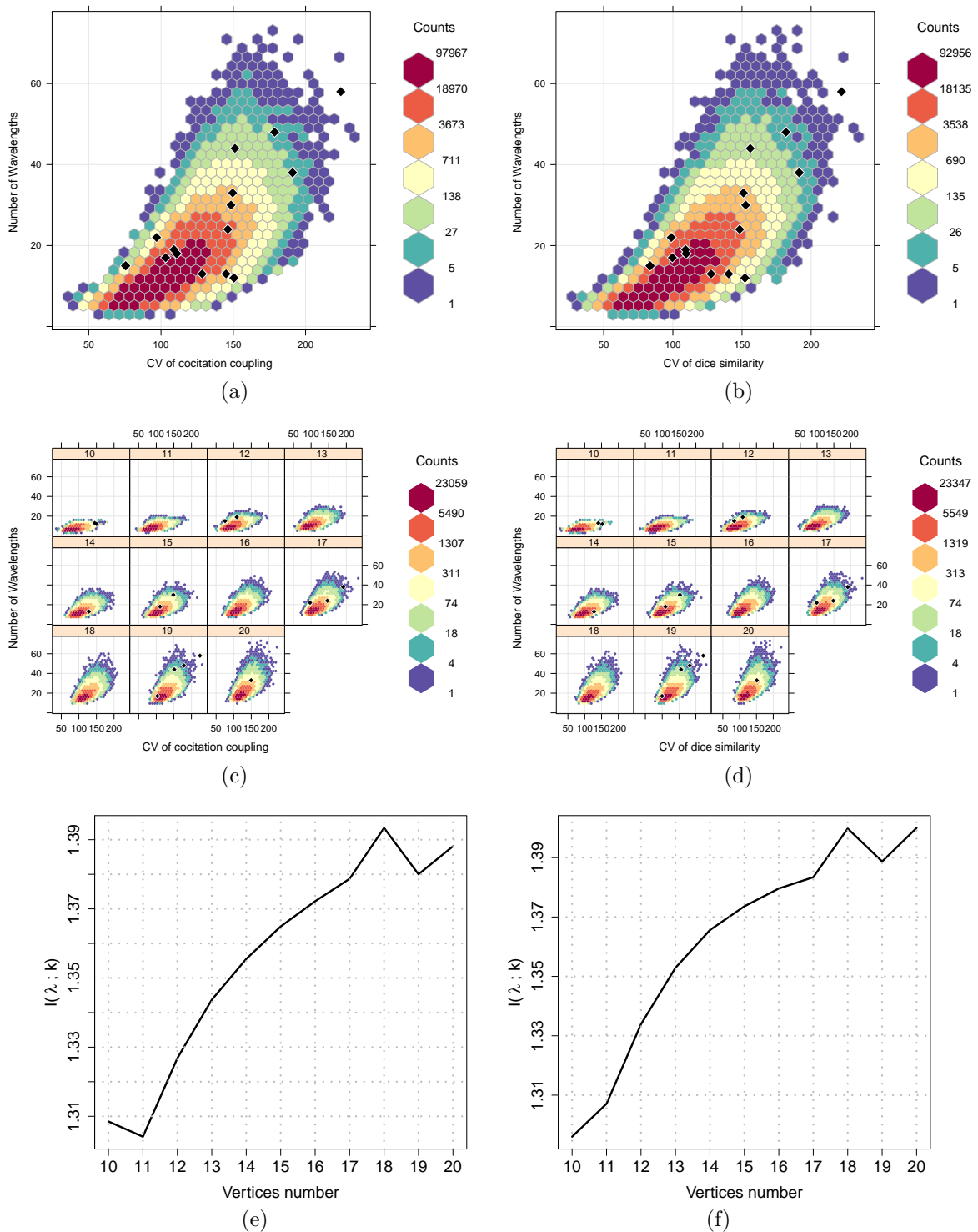


Figura 43 – Coefficient of variation of cocitation coupling ($k = 33$) and Coefficient of variation of dice similarity ($k = 34$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of cocitation coupling and Coefficient of variation of dice similarity both with λ , as a function of network order.

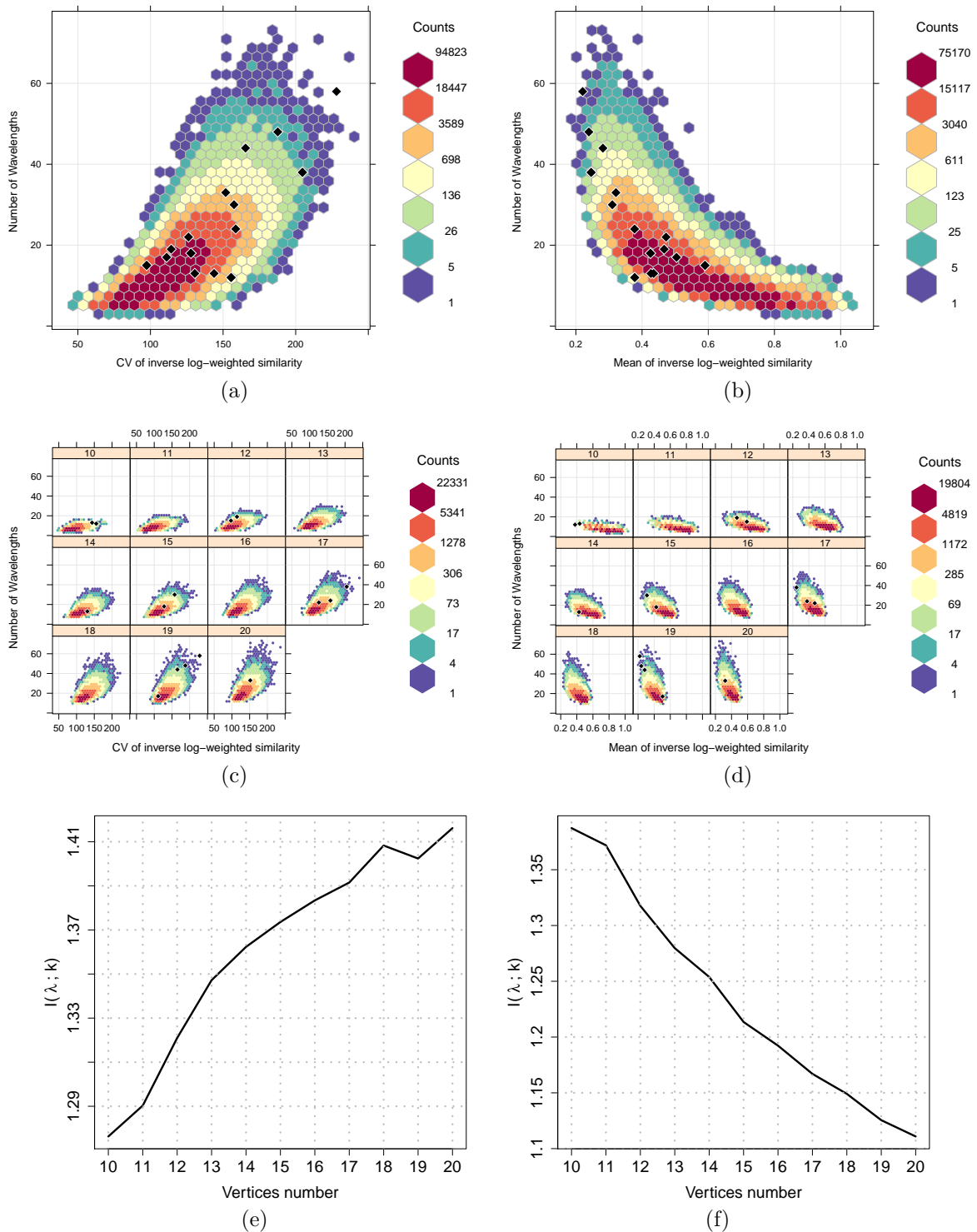


Figura 44 – Coefficient of variation of inverse log-weighted similarity ($k = 35$) and Mean of inverse log-weighted similarity ($k = 36$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of inverse log-weighted similarity and Mean of inverse log-weighted similarity both with λ , as a function of network order.

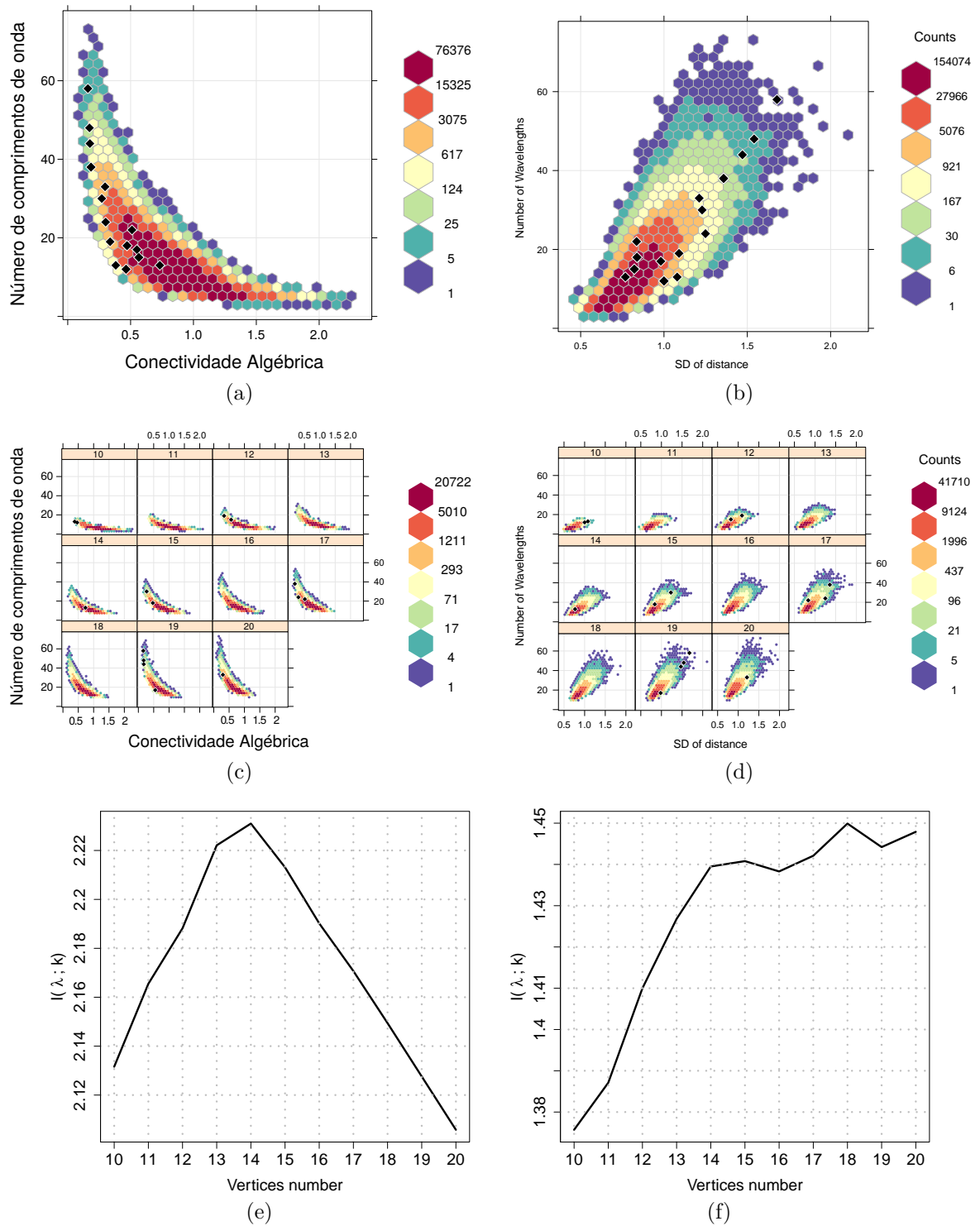


Figura 45 – Algebraic connectivity ($k = 40$) and Standard deviation of distance ($k = 45$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Algebraic connectivity and Standard deviation of distance both with λ , as a function of network order.

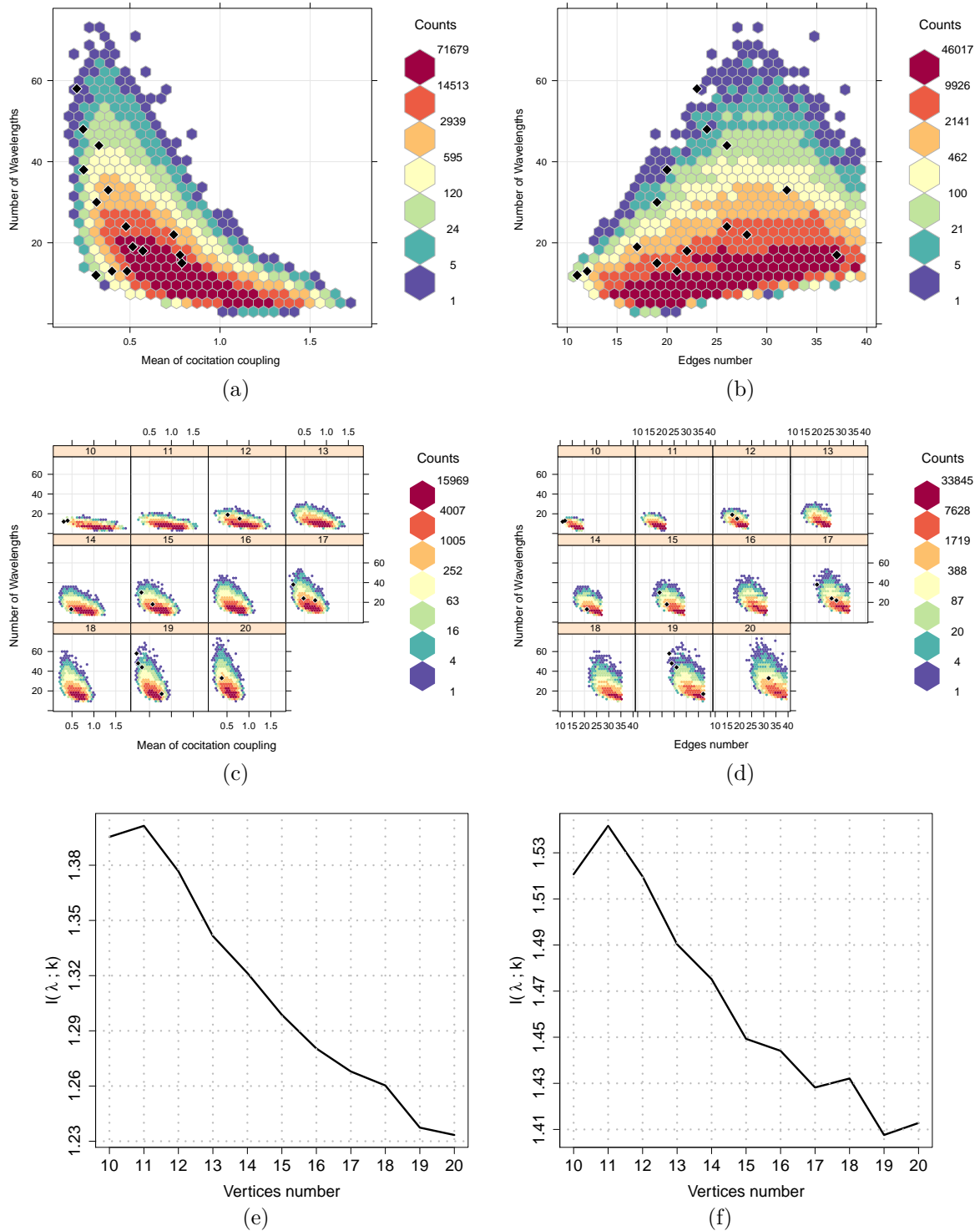


Figura 46 – Mean of cocitation coupling ($k = 46$) and Edges number ($k = 52$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean of cocitation coupling and Edges number both with λ , as a function of network order.

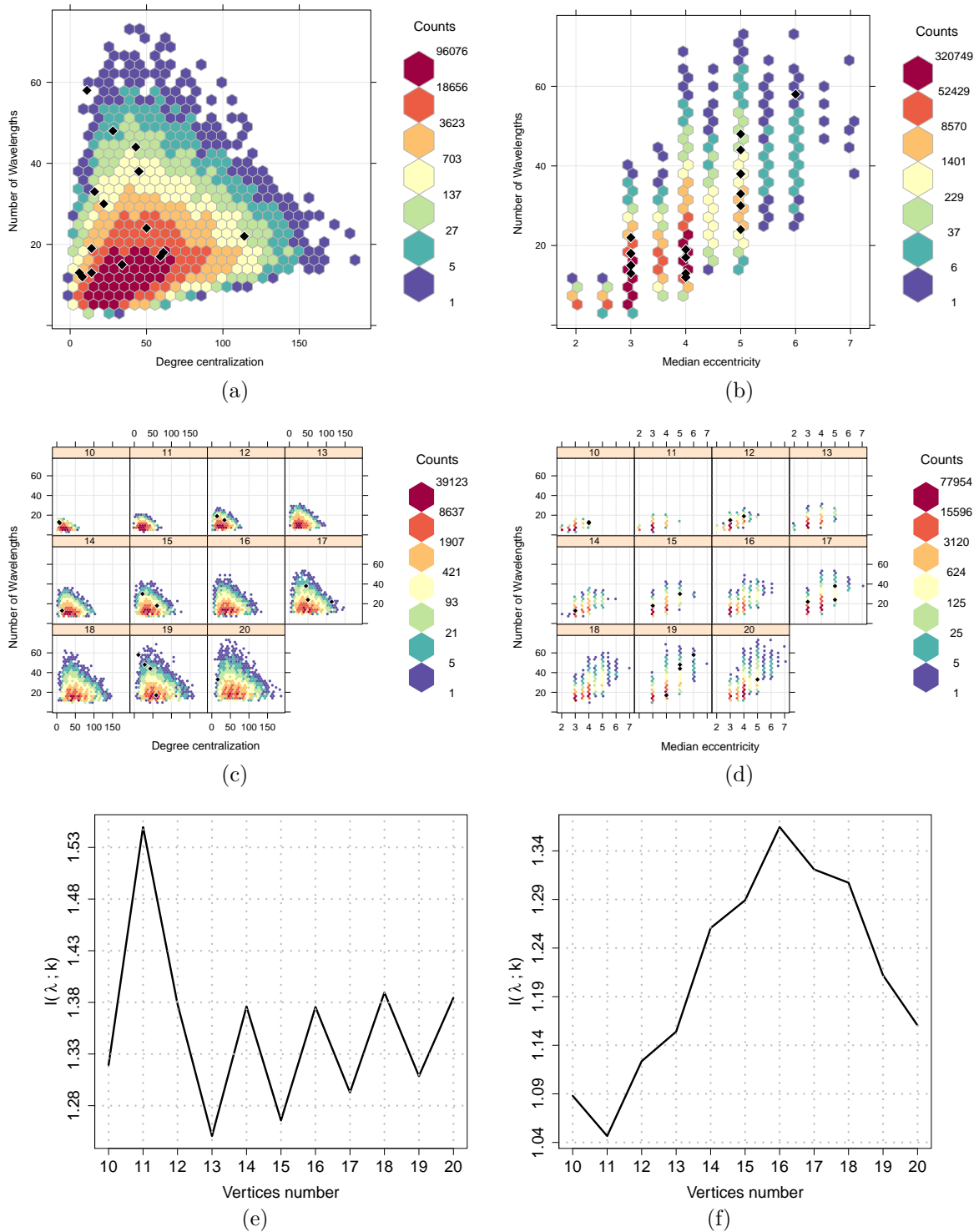


Figura 47 – Degree centralization ($k = 53$) and Median eccentricity ($k = 55$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Degree centralization and Median eccentricity both with λ , as a function of network order.

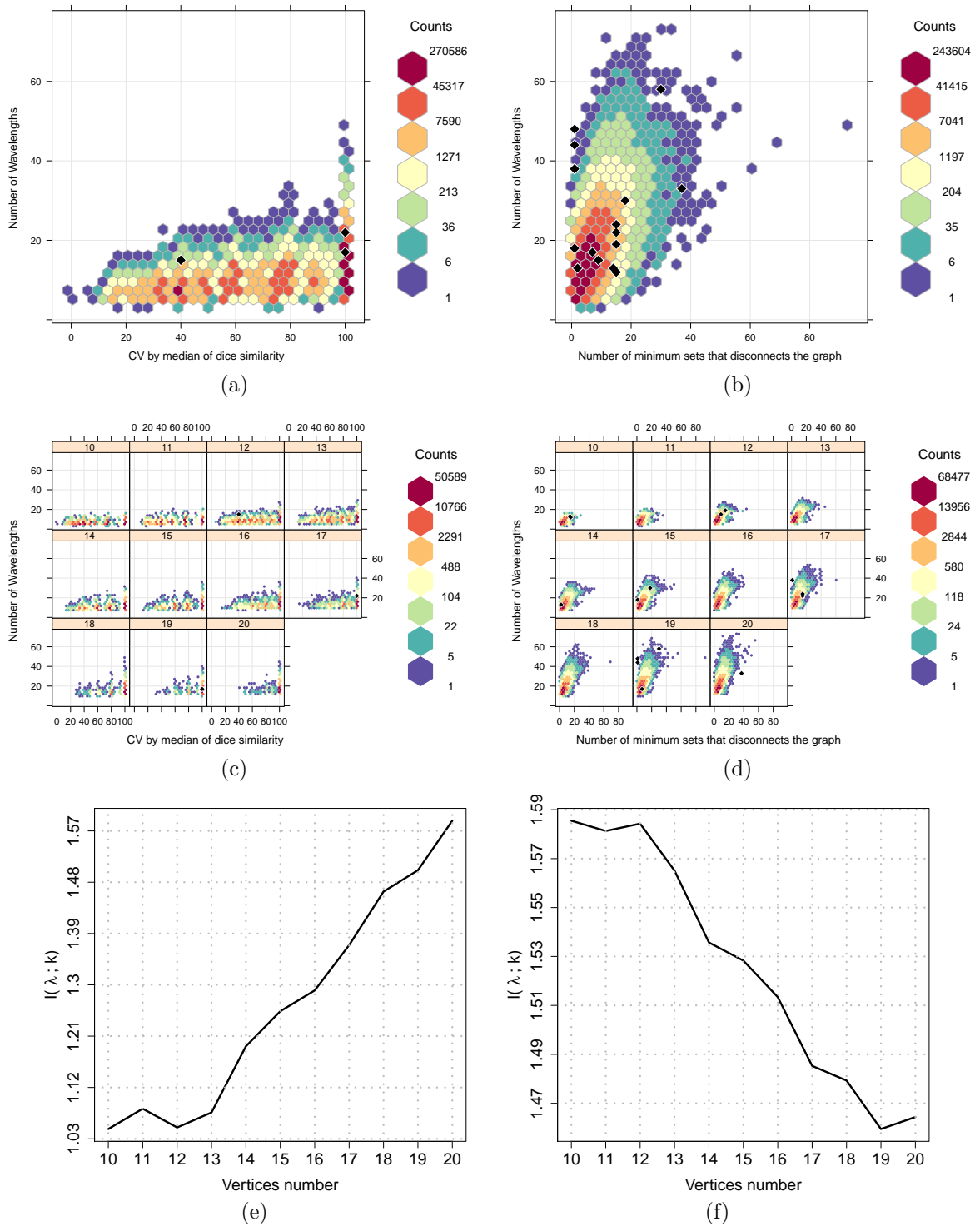


Figura 48 – Coefficient of variation by median of dice similarity ($k = 57$) and Number of minimum sets that disconnects the graph ($k = 59$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation by median of dice similarity and Number of minimum sets that disconnects the graph both with λ , as a function of network order.

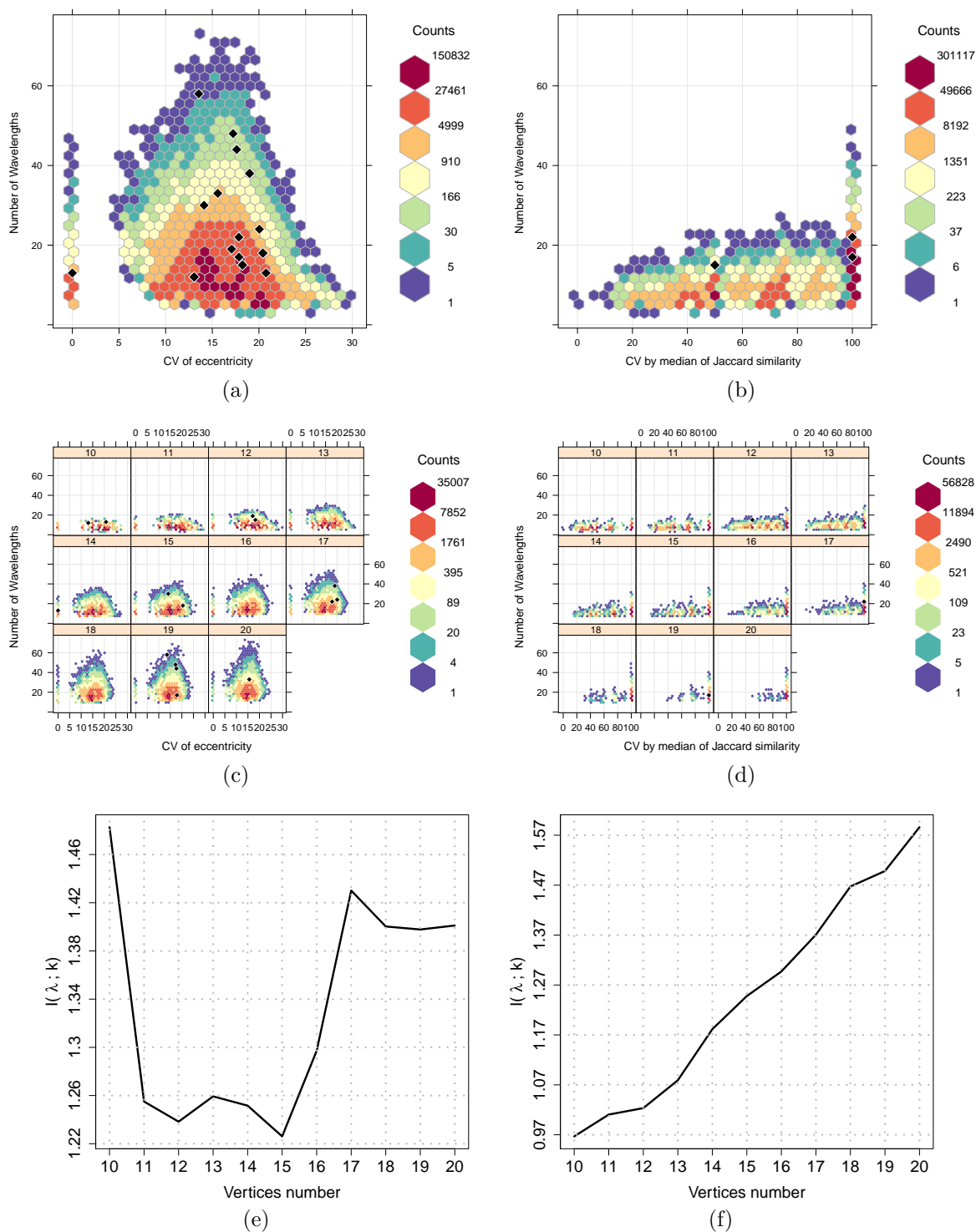


Figura 49 – Coefficient of variation of eccentricity ($k = 60$) and Coefficient of variation by median of Jaccard similarity ($k = 61$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of eccentricity and Coefficient of variation by median of Jaccard similarity both with λ , as a function of network order.

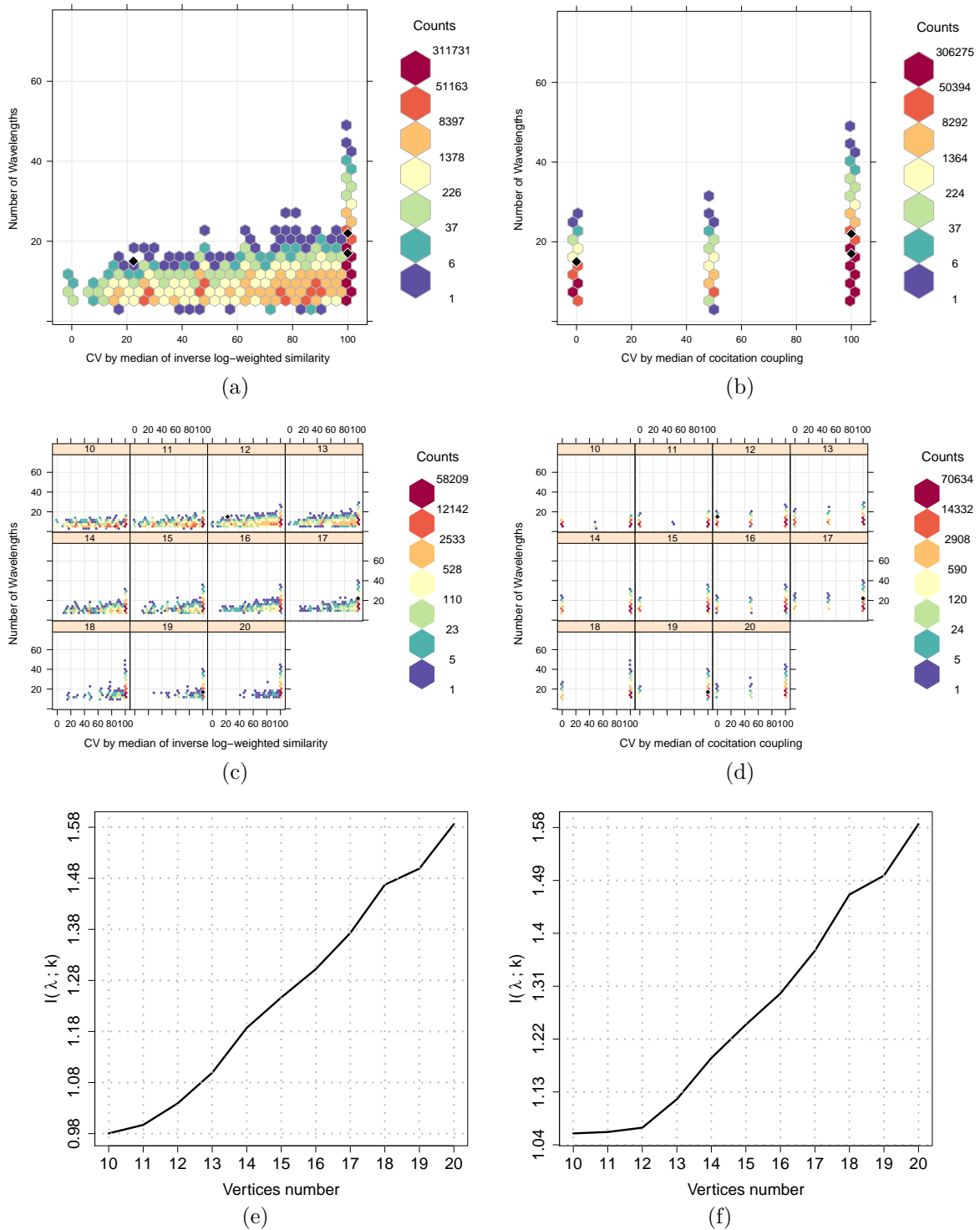


Figura 50 – Coefficient of variation by median of inverse log-weighted similarity ($k = 63$) and Coefficient of variation by median of cocitation coupling ($k = 68$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation by median of inverse log-weighted similarity and Coefficient of variation by median of cocitation coupling both with λ , as a function of network order.

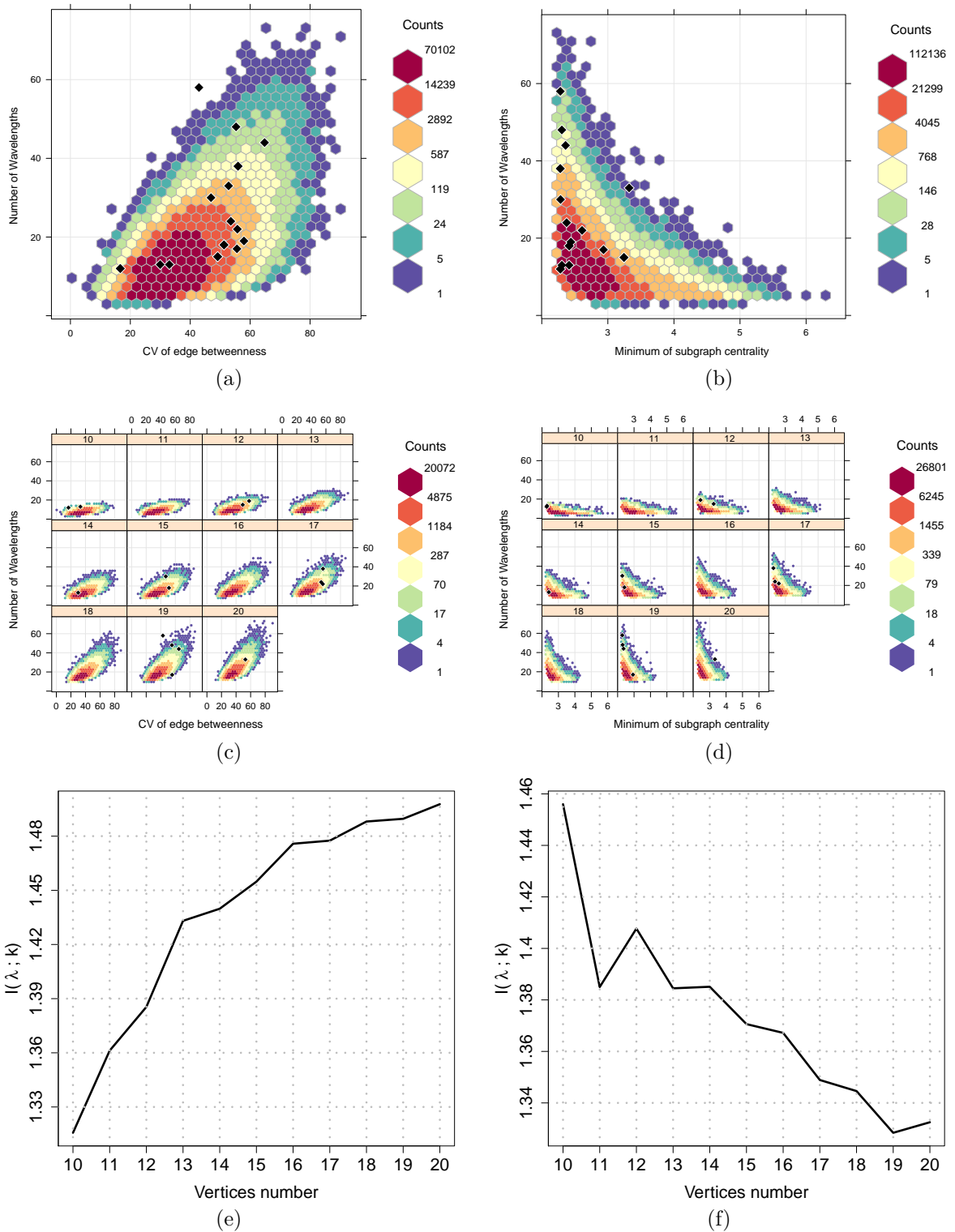


Figure 51 – Coefficient of variation of edge betweenness ($k = 93$) and Minimum of subgraph centrality ($k = 96$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of edge betweenness and Minimum of subgraph centrality both with λ , as a function of network order.

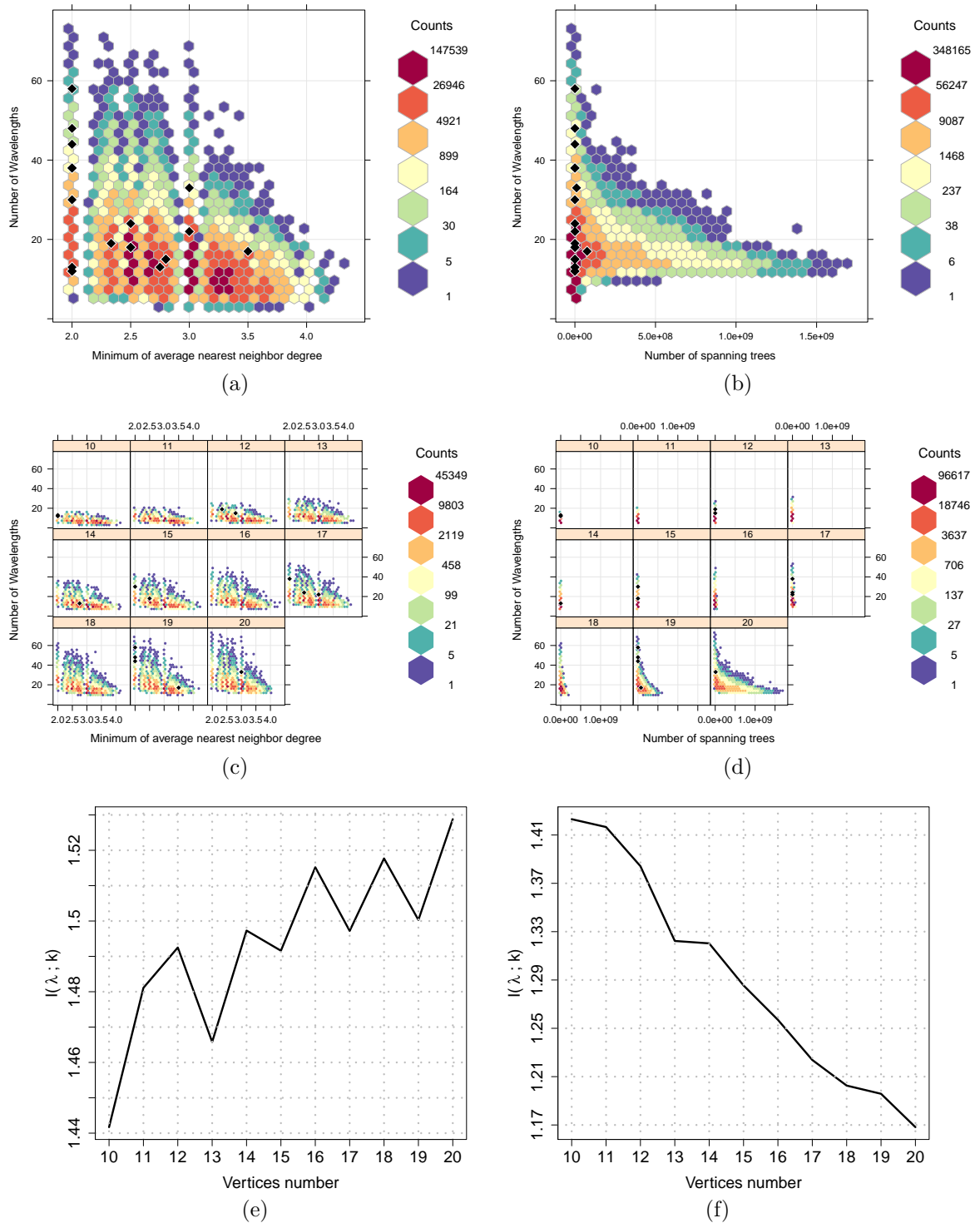


Figura 52 – Minimum of average nearest neighbor degree ($k = 101$) and Number of spanning trees ($k = 109$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Minimum of average nearest neighbor degree and Number of spanning trees both with λ , as a function of network order.

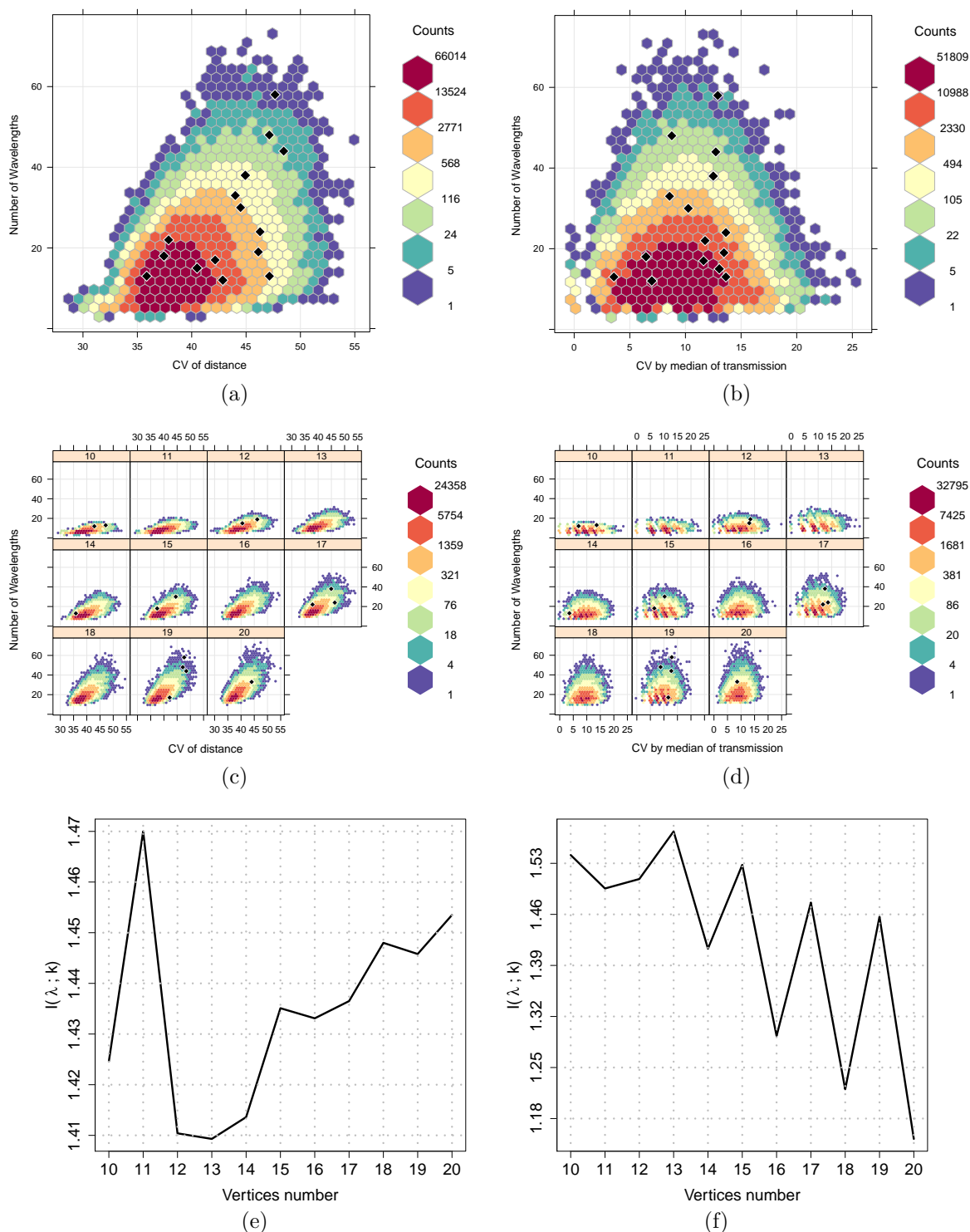


Figura 53 – Coefficient of variation of distance ($k = 115$) and Coefficient of variation by median of transmission ($k = 126$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of distance and Coefficient of variation by median of transmission both with λ , as a function of network order.

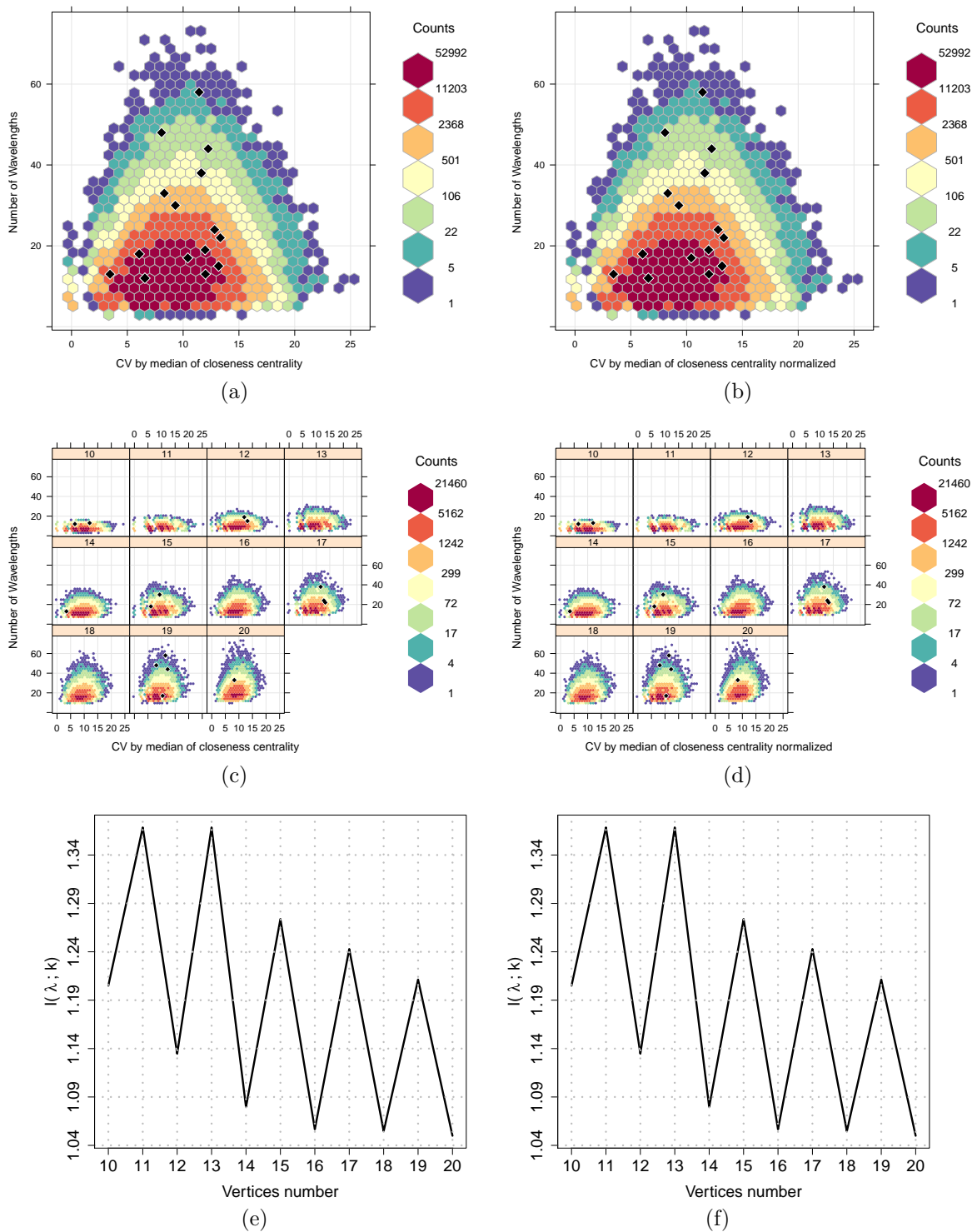


Figura 54 – Coefficient of variation by median of closeness centrality ($k = 138$) and Coefficient of variation by median of closeness centrality normalized ($k = 139$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation by median of closeness centrality and Coefficient of variation by median of closeness centrality normalized both with λ , as a function of network order.

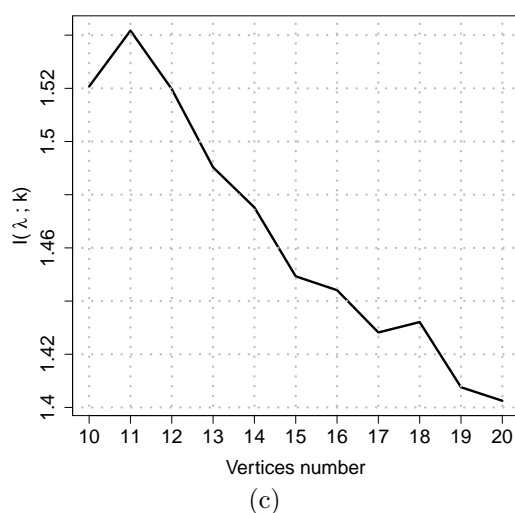
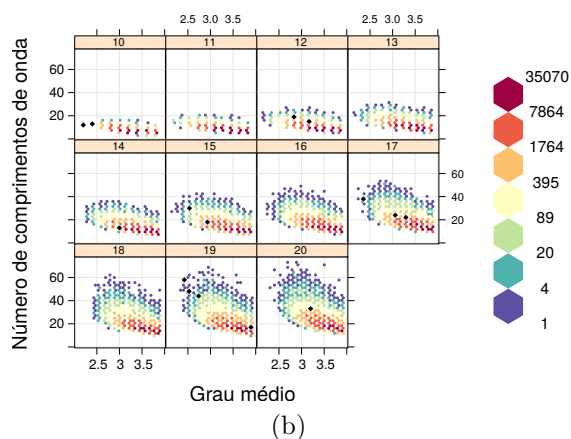
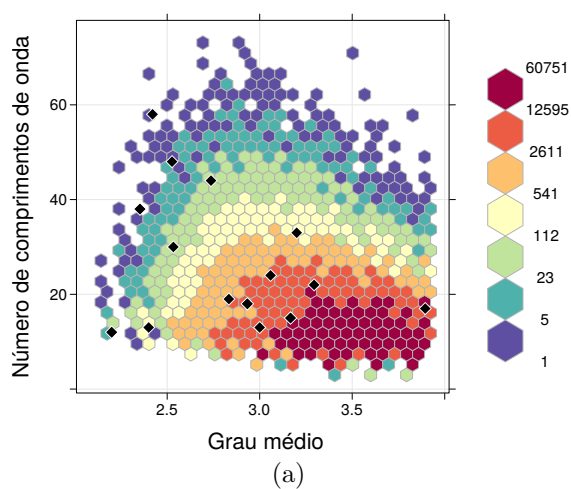


Figura 55 – Mean of vertex degree ($k = 167$) versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a), and separated by n in (b). In (c), mutual information of Mean of vertex degree and λ , as a function of network order.

APÊNDICE E – Tamanho da amostra em modelos lineares generalizados

Considerando o caso em que se deseje executar uma regressão linear com mais de uma, suponhamos j , variáveis explicativas (X_1, X_2, \dots, X_j) para estimar uma variável resposta Y , tem-se aí uma regressão múltipla. E caso a distribuição de probabilidade da variável resposta Y não ser a distribuição normal, tem-se o caso chamado na literatura de Modelos Lineares Generalizados (do inglês, *Generalized Linear Models* - GLM) (FOX, 2015).

Caso haja correlação entre as variáveis explicativas e a variável resposta, o tamanho de amostra para tornar significativa uma regressão múltipla pode ser obtido via o poder do teste em Modelos Lineares Generalizados (COHEN, 1987). Testa-se a regressão com uma hipótese nula de que nenhuma das variáveis explicativas X_1, X_2, \dots, X_j explica a variável resposta Y . Isso significaria que seus coeficientes de regressão são estatisticamente indistinguíveis de 0, logo não existiria regressão. A hipótese alternativa é que pelo menos um dos coeficientes das variáveis explicativas X_1, X_2, \dots, X_j não vale 0, logo existiria uma regressão definida. Isto é testado com um teste-F.

O teste-F tem graus de liberdade no numerador e no denominador. O grau de liberdade do numerador, u , é o número de coeficientes que se terá no modelo (a menos do intercepto). No presente caso, $u = 4$ (como é dado na Seção 3.1). O grau de liberdade do denominador, v , é o número de graus de liberdade do erro: $v = N - u - 1$. Isso implica em $N = v + u + 1$. Assim, como já temos u , obtendo v obtemos N , que é o tamanho de amostra mínimo tal que permite rejeitar a hipótese nula e tornar significativa a regressão quando ela realmente for. E para obter v precisa-se, além do valor de u , do **tamanho do efeito**, do **nível de significância**, e do **poder do teste**.

O **tamanho do efeito**, f^2 , é dado por $R^2/(1 - R^2)$, onde R^2 é o coeficiente de determinação, também conhecido como a “proporção de variância explicada”. Para determinar o tamanho do efeito, supõe-se a proporção de variância que o modelo explica, o R^2 . Especificar um tamanho de efeito pode parecer uma tarefa complicada. Existem muitas sugestões de fórmulas na literatura, até do próprio Cohen (1987), que devem ser vistas apenas como diretrizes muito grosseiras. A própria experiência do profissional no assunto deve ser mais levada em conta. Por exemplo, sendo bem pessimista, supõe-se que o modelo em questão explique 1% da variação na variável resposta, e com isso o tamanho do efeito é $f^2 = 0,01/(1 - 0,01) \approx 0,01$. O pior que pode ocorrer de uma suposição pessimista é o aumento do tamanho da amostra N .

O **nível de significância**, ou erro tipo I, é a probabilidade de rejeitar a hipótese

nula, se ela for verdadeira. A chamada confiança estatística do teste é dada por: (1 - Erro Tipo I). Sendo também exigente, considera-se aqui um erro tipo I de 1%, logo uma confiança de 99%.

Já o erro tipo II é a probabilidade de não rejeitar a hipótese nula quando ela é falsa. O **poder do teste** é dado por: (1 - Erro Tipo II). Sendo também exigente, considera-se aqui um erro tipo II de 2%, logo um poder de 98%.

Ser exigente com a confiança ou com poder do teste apenas acarretam também num aumento final do tamanho da amostra N .

De posse desses valores, [Cohen \(1987\)](#) propôs uma formulação para calcular o grau de liberdade do denominador, v , a qual é possível executar no programa R por meio da função `pwr.f2.test`. Feita essa execução, a seguinte saída é liberada:

```
> pwr.f2.test(u = 4, f2 = 0.01/(1 - 0.01), sig.level = 0.01,  
power = 0.98)
```

```
Multiple regression power calculation
```

```
u = 4
```

```
v = 2842.417
```

```
f2 = 0.01010101
```

```
sig.level = 0.01
```

```
power = 0.98
```

Logo, vê-se que $v \approx 2842$, e como $N = v + u + 1$, temos que $N = 2842 + 4 + 1$, resultando em $N = 2847$.

APÊNDICE F – Aplicação das heurísticas nas redes reais

OBS.: G_S é o grafo solução fornecido pela heurística.

Redes Reais	A = λ exato da rede original	B = $E[\lambda]$ de G_S	C = λ exato de G_S	Diferença entre estimativa e obtido: C – B (abs)	Diferença entre estimativa e obtido: C – B (em %)	Melhora Total A – C (abs)	Melhora Total A – C (em %)
BUSCA ALEATÓRIA							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	7	7	0	0%	8	53%
NSFNET	13	10	10	0	0%	3	23%
AUSTRIA	18	11	12	1	8%	6	33%
MZIMA	30	15	15	0	0%	15	50%
ARNES	38	21	23	2	9%	15	39%
GERMANY	24	12	13	1	8%	11	46%
SPAIN	22	11	12	1	8%	10	45%
MEMOREX	48	23	24	1	4%	24	50%
CANARIE	44	19	19	0	0%	25	57%
EON	17	10	11	1	9%	6	35%
ARPANET	33	16	16	0	0%	17	52%
DOWNHILL (GULOSO) COM <i>sorted</i> = FALSE							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	7	8	1	13%	7	47%
NSFNET	13	9	10	1	10%	3	23%
AUSTRIA	18	11	12	1	8%	6	33%
MZIMA	30	14	16	2	13%	14	47%
ARNES	38	20	22	2	9%	16	42%
GERMANY	24	12	13	1	8%	11	46%
SPAIN	22	11	13	2	15%	9	41%
MEMOREX	48	22	22	0	0%	26	54%
CANARIE	44	18	18	0	0%	26	59%
EON	17	9	11	2	18%	6	35%
ARPANET	33	15	16	1	6%	17	52%

OBS.: G_S é o grafo solução fornecido pela heurística.

Redes Reais	A = λ exato da rede original	B = $E[\lambda]$ de G_S	C = λ exato de G_S	Diferença entre estimativa e obtido: C - B (abs)	Diferença entre estimativa e obtido: C - B (em %)	Melhora Total A - C (abs)	Melhora Total A - C (em %)
DOWNHILL (GULOSO) COM <i>sorted</i> = TRUE							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	8	8	0	0%	7	47%
NSFNET	13	12	13	1	8%	0	0%
AUSTRIA	18	12	13	1	8%	5	28%
MZIMA	30	15	17	2	12%	13	43%
ARNES	38	22	24	2	8%	14	37%
GERMANY	24	13	14	1	7%	10	42%
SPAIN	22	12	14	2	14%	8	36%
MEMOREX	48	24	25	1	4%	23	48%
CANARIE	44	21	21	0	0%	23	52%
EON	17	16	16	0	0%	1	6%
ARPANET	33	33	33	0	0%	0	0%
BUSCA LOCAL COM <i>set</i> = S							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	6	7	1	14%	8	53%
NSFNET	13	9	10	1	10%	3	23%
AUSTRIA	18	11	12	1	8%	6	33%
MZIMA	30	14	16	2	13%	14	47%
ARNES	38	20	22	2	9%	16	42%
GERMANY	24	12	14	2	14%	10	42%
SPAIN	22	10	11	1	9%	11	50%
MEMOREX	48	21	21	0	0%	27	56%
CANARIE	44	17	18	1	6%	26	59%
EON	17	9	10	1	10%	7	41%
ARPANET	33	13	15	2	13%	18	55%

OBS.: G_S é o grafo solução fornecido pela heurística.

Redes Reais	A = λ exato da rede original	B = $E[\lambda]$ de G_S	C = λ exato de G_S	Diferença entre estimativa e obtido: C - B (abs)	Diferença entre estimativa e obtido: C - B (em %)	Melhora Total A - C (abs)	Melhora Total A - C (em %)
BUSCA LOCAL COM $set = M$							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	6	7	1	14%	8	53%
NSFNET	13	9	10	1	10%	3	23%
AUSTRIA	18	10	11	1	9%	7	39%
MZIMA	30	14	16	2	13%	14	47%
ARNES	38	20	22	2	9%	16	42%
GERMANY	24	12	13	1	8%	11	46%
SPAIN	22	10	11	1	9%	11	50%
MEMOREX	48	20	21	1	5%	27	56%
CANARIE	44	18	19	1	5%	25	57%
EON	17	9	10	1	10%	7	41%
ARPANET	33	14	15	1	7%	18	55%
BUSCA LOCAL COM $set = G$							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	6	7	1	14%	8	53%
NSFNET	13	9	10	1	10%	3	23%
AUSTRIA	18	11	12	1	8%	6	33%
MZIMA	30	14	15	1	7%	15	50%
ARNES	38	20	22	2	9%	16	42%
GERMANY	24	13	14	1	7%	10	42%
SPAIN	22	11	12	1	8%	10	45%
MEMOREX	48	21	22	1	5%	26	54%
CANARIE	44	18	19	1	5%	25	57%
EON	17	14	13	1	8%	4	24%
ARPANET	33	19	20	1	5%	13	39%

OBS.: G_S é o grafo solução fornecido pela heurística.

Redes Reais	A = λ exato da rede original	B = $E[\lambda]$ de G_S	C = λ exato de G_S	Diferença entre estimativa e obtido: C - B (abs)	Diferença entre estimativa e obtido: C - B (em %)	Melhora Total A - C (abs)	Melhora Total A - C (em %)
VNS COM $N_{type} = Edges$							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	7	8	1	13%	7	47%
NSFNET	13	9	10	1	10%	3	23%
AUSTRIA	18	11	12	1	8%	6	33%
MZIMA	30	14	15	1	7%	15	50%
ARNES	38	21	23	2	9%	15	39%
GERMANY	24	12	13	1	8%	11	46%
SPAIN	22	10	12	2	17%	10	45%
MEMOREX	48	22	22	0	0%	26	54%
CANARIE	44	18	19	1	5%	25	57%
EON	17	8	10	2	20%	7	41%
ARPANET	33	14	15	1	7%	18	55%
VNS COM $N_{type} = KDS$							
VBNS	19	8	9	1	11%	10	53%
CESNET	15	7	8	1	13%	7	47%
NSFNET	13	9	10	1	10%	3	23%
AUSTRIA	18	10	11	1	9%	7	39%
MZIMA	30	14	15	1	7%	15	50%
ARNES	38	20	22	2	9%	16	42%
GERMANY	24	12	13	1	8%	11	46%
SPAIN	22	10	11	1	9%	11	50%
MEMOREX	48	20	21	1	5%	27	56%
CANARIE	44	17	18	1	6%	26	59%
EON	17	9	11	2	18%	6	35%
ARPANET	33	13	15	2	13%	18	55%

APÊNDICE G – Redes reais e redes soluções fornecidas pela heurística VNS-KDS

LISTA DE ARESTAS DAS REDES REAIS DE ENTRADA DAS HEURÍSTICAS:
(ordenadas pela ordem n)

VBNS 3 10 5 7 11 12 2 4 1 10 9 12 4 11 6 8 7 10 7 8 3 9 2 3 4 6
6 9 7 12 1 4 2 5

CESNET 1 11 5 7 6 7 7 8 7 9 9 10 10 11 11 12 4 12 3 7 1 3 2 4 3
4 2 6 8 11 4 9 5 11 3 9 2 11

NSFNET 1 14 4 6 6 7 6 14 7 8 11 12 2 5 8 10 4 9 2 3 9 12 11 13
10 14 1 4 3 8 5 11 3 12 1 2 8 9 6 13 5 10

AUSTRIA 1 3 3 4 2 6 5 8 6 8 3 9 4 11 2 12 5 12 7 12 6 13 10 15
14 15 12 14 9 15 8 10 3 7 8 11 3 13 11 14 1 15 2 9

MZIMA 4 7 6 7 3 8 9 12 2 4 9 15 5 13 3 10 6 13 2 11 2 12 5 15 7
14 2 10 9 14 8 9 1 13 1 3 11 15

ARNES 2 6 4 8 4 9 1 10 5 11 5 12 5 14 13 14 12 16 3 17 4 7 10
13 1 17 7 13 6 15 2 17 3 8 9 15 6 11 3 16

GERMANY 1 17 2 3 2 16 4 5 7 8 13 16 5 9 4 10 12 16 4 12 1 16 3
15 3 10 8 11 15 17 6 11 4 8 2 11 12 14 5 15 5 6 13 15 7 16
11 14 9 16 11 17

SPAIN 1 2 1 8 1 9 4 9 6 9 7 10 8 9 9 17 9 13 9 15 10 11 12 14 7
12 9 11 7 16 9 16 2 12 5 12 6 17 12 15 3 9 4 10 4 14 5 10
6 12 7 13 1 10 3 4

MEMOREX 8 10 14 15 15 16 14 18 13 19 3 17 2 19 4 16 11 12 7 18
13 17 5 9 4 12 1 19 2 7 8 14 1 16 2 11 10 12 3 10 5 6 9 18
6 16 5 17

CANARIE 3 4 1 8 5 8 8 9 6 10 7 13 13 15 10 18 12 18 14 19 1 12
13 17 6 14 12 15 2 13 5 16 4 7 4 14 8 19 4 11 9 13 16 17 3 5
11 12 2 6 16 18

EON 1 17 2 3 2 8 2 9 2 19 3 4 4 6 5 6 6 11 6 15 7 8 9 17 10 16
11 12 14 15 15 17 17 18 6 13 2 14 5 10 4 15 3 8 8 11 4 16 1 6
8 16 10 17 9 12 10 19 2 13 12 14 6 12 7 15 10 15 9 16 2 18 7 17

ARPANET 1 2 1 17 1 18 2 3 2 12 3 4 3 11 4 5 5 9 6 10 7 8 8 9 9
13 10 15 13 14 14 20 15 16 15 20 16 17 6 19 7 18 11 14 9 17
10 18 11 16 7 19 8 14 12 13 5 10 6 12 4 19 6 20

LISTA DE ARESTAS DAS REDES SOLUÇÕES DE SAÍDA DA HEURÍSTICA VNS-KDS (as quais diferem das redes de entrada apenas por um rearranjo de arestas):

VBNS 7 6 4 3 10 2 8 9 5 6 11 2 9 10 12 1 3 6 3 1 7 11 8 7 9 12
12 11 3 2 5 9 8 4

CESNET 6 12 10 11 2 11 11 1 11 8 8 9 9 12 12 5 4 5 7 11 6 7 3 4
7 4 3 2 1 12 4 8 10 12 7 8 3 12

NSFNET 4 5 3 6 6 2 6 5 2 14 10 11 8 9 14 7 3 13 8 12 13 11 10 1
7 5 4 3 12 14 9 10 12 11 4 8 14 13 6 1 9 7

AUSTRIA 8 12 12 1 14 15 4 5 15 5 12 11 1 2 14 9 4 9 10 9 15 13
6 7 3 7 9 3 11 7 5 6 12 10 5 2 12 13 2 3 8 7 14 11

MZIMA 5 4 2 4 15 13 10 7 6 5 10 9 11 12 15 1 2 12 6 8 6 7 11 9
4 3 6 1 10 3 13 10 14 12 14 15 8 9

ARNES 4 3 17 16 17 1 6 8 12 13 12 7 12 11 10 11 7 14 15 5 17 9
8 10 6 5 9 10 3 2 4 5 15 16 1 2 3 13 15 14

GERMANY 1 17 13 10 13 14 5 8 3 4 12 14 8 9 5 6 15 14 5 15 1 14
10 11 10 6 4 16 11 17 7 16 5 4 13 16 15 2 8 11 8 7 12 11 3 14
16 2 9 14 16 17

SPAIN 5 7 5 4 5 8 6 8 2 8 12 17 4 8 8 1 8 15 8 9 17 13 10 16 12
10 8 13 12 14 8 14 7 10 11 10 2 1 10 9 3 8 6 17 6 16 11 17 2 10
12 15 5 17 3 6

MEMOREX 7 19 10 18 18 16 10 8 14 17 11 13 3 17 4 16 1 2 6 8 14
13 12 9 4 2 5 17 3 6 7 10 5 16 3 1 19 2 11 19 12 15 9 8 15 16
12 13

CANARIE 1 19 6 5 4 5 5 16 11 12 17 8 8 15 12 10 7 10 2 3 6 7 8
14 11 2 7 15 13 8 4 9 19 17 19 2 5 3 19 18 16 8 9 14 1 4 18 7
13 11 9 10

EON 18 2 10 17 10 8 10 6 10 5 17 16 16 15 4 15 15 14 15 9 1 8 6
2 3 7 14 12 11 9 9 2 2 19 15 13 10 11 4 3 16 9 17 8 8 14 16 7
18 15 8 7 3 2 6 12 3 5 10 13 12 11 15 12 1 9 3 9 6 7 10 19 1 2

ARPANET 19 18 19 20 19 10 18 13 18 1 13 8 13 14 8 7 7 16 4 9 5
2 2 16 16 17 9 11 17 12 12 3 11 15 11 3 15 20 4 6 5 10 14 12 16
20 9 10 14 15 5 6 2 12 1 17 7 9 4 1 8 6 4 3

Anexos

ANEXO A – Evaluating Community Detection Methods in a Controlled Experiment

Abstract: Detecting communities in non-trivial networks is a hard task that has been attracting the attention of many researchers in the field of complex networks. When the network has identical and non-overlapping communities, however, one could expect this task to be easy. We show in this paper that, surprisingly, it is not the case. For this purpose, we carried out a controlled experiment in which 9 well-known community detection methods are applied to 8 different hierarchical ring networks of 60 nodes each. None of the tested methods correctly identified the communities for all networks. Moreover, the modularity score showed a weak response in the identification of rings as communities in hierarchical ring networks. These results point to the need of discussing what community detection algorithms are actually detecting, and how to evaluate their results, both when dealing with trivial or non-trivial networks.

Keywords: *questioning community detection methods, hierarchical ring networks, modularity index, graph theory*

A.1 Introduction

In the present study, the considered concept of community is given by a group of elements with more connections between their members than with members of other communities. In some applications, when using community detection methods in networks, the results obtained may do not match what is expected. In Figure 56 is displayed an application of the Edge betweenness method (NEWMAN; GIRVAN, 2004) for community detection. In the case of the network of Figure 56a, formed by joining 3 complete graphs of 20 nodes each, is shown a coherent and expected detection of communities with a modularity value (index proposed in (CLAUSET; NEWMAN; MOORE, 2004) and widely used to measure how good a division of communities is) given by 0.66. In the network of Figure 56b, formed by the junction of 3 cycles of 20 nodes each, is displayed that the method could not detect the 3 cycles. Such detection did not occur, but it was expected to occur by the very concept of community, that is even weakly interconnected, there are more connections within the cycles than between them. It is also surprising that the value of modularity in the division of communities in the network in Figure 56b is 0.69, a larger value than that of the network with coherent detection of communities from Figure 56a.

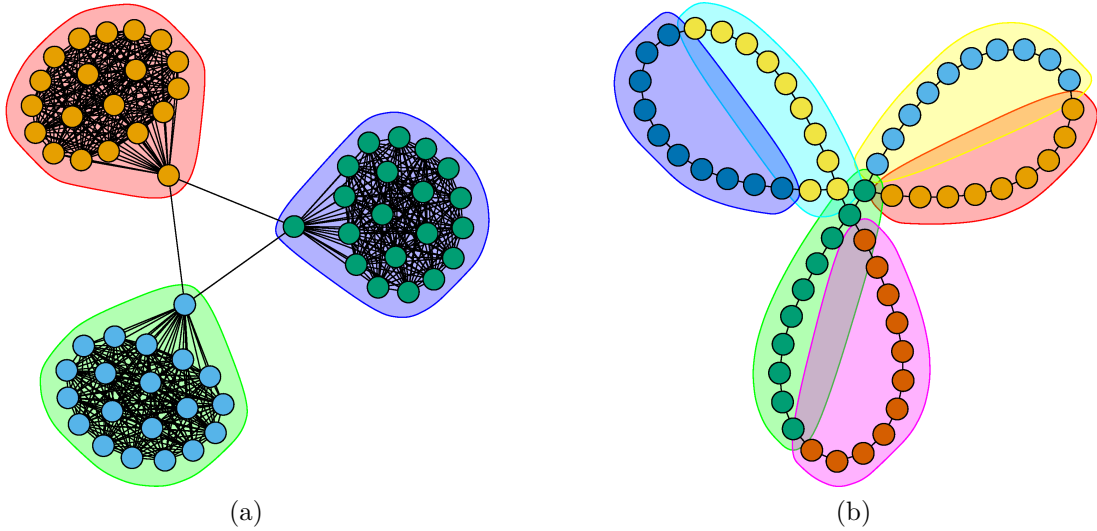


Figura 56 – Network (a) is formed by joining 3 complete graphs of 20 nodes each, and network (b) is formed by joining 3 cycles of 20 nodes each. In both networks the Edge betweenness method is used for the division into communities. In (a), the modularity value is 0.66, and in (b) is 0.69, even (a) presenting greater visual structuring of communities than (b).

The first optical backbone networks were created in the world in the beginning of the second half of the last century, and since then cycles, usually called ring networks, are among the first options when thinking about the topology to be used (SIMMONS, 2014). Hierarchical self-healing ring networks (SHI; FONSEKA, 1995) came as the next step to optical networks in rings, in which there is a top-level central ring, called level 0. Each of the other rings that bind to the ring of level 0, through a single node, are called level-1 rings, as shown in Figure 57.

In hierarchical ring networks, it is clear the perception of the rings, of level 0 or 1, as community network structures. It is questioned how methods of community detection would respond in the identification of such rings, since, to the human eye, they are easily detectable structures.

No matter how unrepresentative hierarchical networks may be among graphs – fixed a number of nodes – what draws attention is the simplicity of the cases they represent, generating an expectation that community detection methods should not have greater difficulties in detecting the rings of hierarchical ring networks: simple community structures should not present challenges to community detection methods consolidated in the literature.

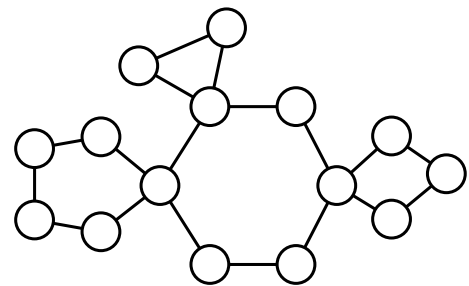


Figura 57 – Example of a hierarchical ring network, with the center ring of level 0 (of 6 nodes) binded to others 3 rings of level 1.

For this, the present study systematically tests, by means of a controlled experiment, nine methods of detecting communities already present in the literature, in a set of hierarchical ring networks, where the detection of communities should be a relatively easy task. This strategy is used to verify the performance of these methods in the detection of communities in the simple case of hierarchical ring networks. The objective is not exactly to identify the rings as community structures, but actually to test the ability of the methods to perform a trivial task, thereby questioning the ability of these methods to detect less trivial community structures in more sophisticated networks.

A.2 Methodology

In summary, the method consists of a comparative analysis of nine community detection methods found in the literature, applied to the simple case of hierarchical ring optical networks.

In order to make the experiment as fair as possible, the hierarchical ring networks are generated as follows: each network has the number of nodes fixed in 60, with level-1 rings of the same order within the network, which means that each networks could have 3, 4, 5, 6, 10, 12, 15 or 20 level-1 rings attached to a central ring of level 0, and consequently the order of each level-1 ring could be given respectively by 20, 15, 12, 10, 6, 5, 4 or 3 nodes, in each network. Thus, a total of 8 networks are analyzed, which can be observed in Figure 58. The networks are modeled by graphs, where the nodes and links of the networks

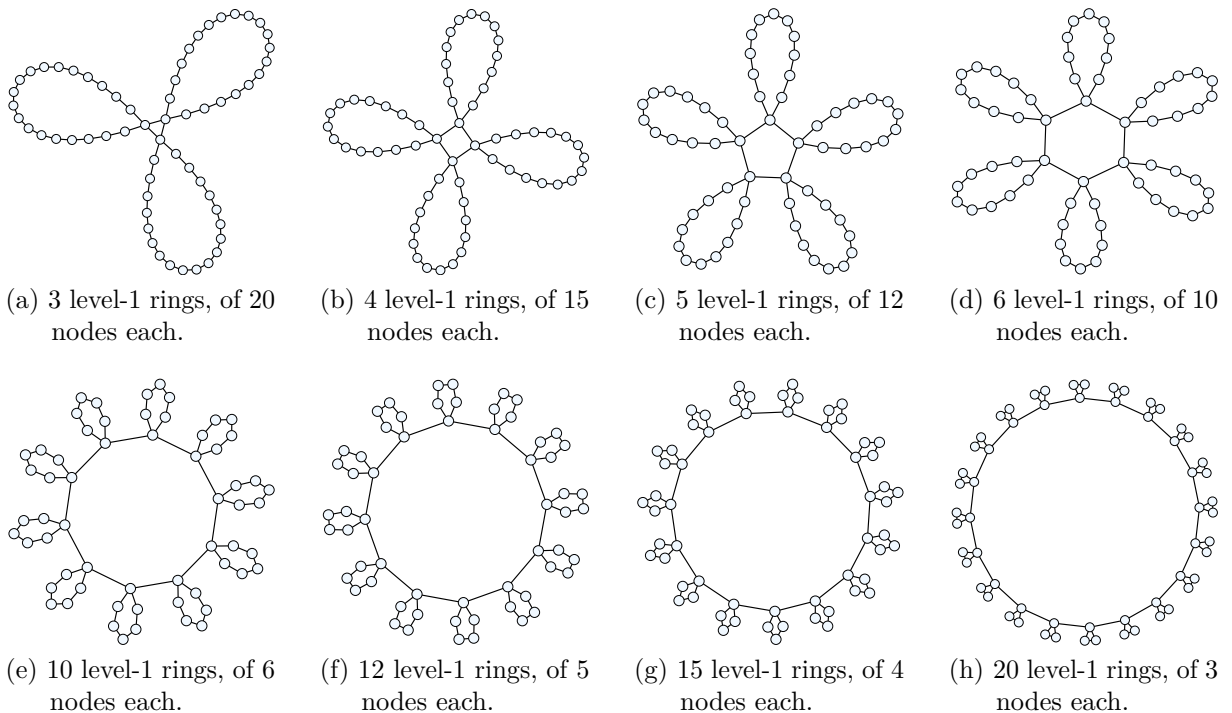


Figure 58 – Hierarchical ring networks of 60 nodes divided homogeneously.

are represented by the vertices and edges of the graphs. These graphs are also considered as non-directed and with unit weight edges.

We test systematically the nine methods of detecting communities in these networks. The choice of the methods is given by accessibility criteria, that is, some of the best known and already implemented methods are chosen, since they are more likely to be used in general studies. Thus, the methods tested are as follows:

- *Edge betweenness*: it is a method for detecting community structures based on edge betweenness, which measures, for each edge, the proportion of the shortest paths passing through that edge (NEWMAN; GIRVAN, 2004).
- *Multilevel modularity*: in this method a multi-level optimization algorithm of the modularity index is implemented to find the community structures (CLAUSET; NEWMAN; MOORE, 2004).
- *Optimal modularity*: it obtains the community structures by maximizing the modularity index of all possible partitions (BRANDES *et al.*, 2008).
- *Walktrap*: it seeks to find communities through random paths. The idea used in this method is that short random paths tend to be in the same community (PONS; LATAPY, 2005).
- *Spin-glass*: it seeks to find community structures using the spin-glass model and simulated annealing (TRAAG; BRUGGEMAN, 2009).
- *Fastgreedy modularity*: this method detects community structures via direct optimization of the modularity index value (CLAUSET; NEWMAN; MOORE, 2004).
- *Infomap*: this method use two nomenclatures, one to distinguish communities in the graph and another to distinguish nodes in a community, and then detects communities minimizing the quantity of information needed to represent any random walk in the graph using this nomenclature (ROSVALL; AXELSSON; BERGSTROM, 2009).
- *Label propagation*: in this method each node is started with a unique label and in each step, each node adopts the label that most of its neighbors have. In this interactive process, densely connected groups of nodes form a consensus of a single label per community (RAGHAVAN; ALBERT; KUMARA, 2007).
- *Eigenvector modularity*: it seeks to find communities by calculating the main non-negative eigenvector of the graph modularity matrix (NEWMAN, 2006).

Each of these methods applies a division of the network into communities according to its own criteria. Given this division, it is possible to calculate the modularity index (CLAUSET; NEWMAN; MOORE, 2004), denoted by Q , which is a parameter that indicates how good the division by communities is. Networks with a high index of modularity have dense connections between nodes within communities, and sparse

connections between nodes of different communities. The modularity index is given by

$$Q = \frac{1}{2m} \sum_{i,j=1}^n \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(i, j), \quad (\text{A.1})$$

where n is the number of network nodes, m is the number of network edges, A is the adjacency matrix, k_i is the degree of the node i , and $\delta(i, j) = 1$, if the nodes i and j are in the same community, and $\delta(i, j) = 0$ otherwise. From this definition, it follows that Q assumes values in the range $[-1/2, 1)$.

The entire implementation is done in the program R (R Core Team, 2016), using the igraph package (CSARDI; NEPUSZ, 2006), which provides all the community detection methods considered in the present analysis¹. All community detection methods have a running time on the order of seconds, on a regular personal computer, except for the Optimal modularity method, which can spend time on the order of a few minutes.

A.3 Results and discussion

A.3.1 Detection of rings as communities in hierarchical ring networks

As a criterion of this study, we say that a method detects communities in a network in an **expected** way, when each level-1 ring in the hierarchical ring network is identified as a community. The **expected** way is also detected when the central ring of level 0 is detected as a community, and the other rings (without the nodes that are in the central ring) form, each one, a community. If adjacent rings are agglutinated two by two, or three by three, for example, we say that the method respond in an **approximate** way, with some coherence but not exactly as expected. Finally, we say that there is **failure**, when the detection of the communities reached by the method does not meet any of the situations defined above. All the 8 hierarchical ring networks under analysis are submitted to all methods of community detection listed in Section A.2. In Table 13, the performance of each method in detecting communities can be observed.

First, it should be noted that in Table 13, all methods have failed to detect rings of the network with 3 level-1 rings, which intuitively appears to be a network of easy detection of their rings as communities. One of these results, for example, is the response given by the Fastgreedy modularity method, which can be seen in Figure 59a. In this figure, we can see that each ring is divided into parts that do not meet expectations.

On the other hand, in Table 13 we can see that the network with 10 level-1 rings is the one that has more methods able to identify its rings as expected. An example of this identification is given by the Edge betweenness method, as shown in Figure 59b.

¹All R codes used in this study are available on the website http://rpubs.com/ddepizzol/evaluating_community_detection_methods.

Tabela 13 – Results of the application of the community detection methods, for each hierarchical ring network analyzed. The symbols ✓ and ✗ indicate, respectively, success and failure in rings detection. An +/- indicates that the detection of the rings is not as expected, but that the result has some coherence.

Communities detection methods	Number of level-1 rings per graph							
	3	4	5	6	10	12	15	20
<i>Edge betweenness</i>	✗	✓	✓	✓	✓	✓	+/-	+/-
<i>Multilevel modularity</i>	✗	✓	✓	✓	✓	✓	+/-	+/-
<i>Optimal modularity</i>	✗	✗	✓	✓	✓	✗	+/-	+/-
<i>Walktrap</i>	✗	✗	✗	✓	✓	✗	+/-	✗
<i>Spin-glass</i>	✗	✗	✗	✓	✓	✗	✗	✗
<i>Fastgreedy modularity</i>	✗	✗	✗	✗	✓	✗	+/-	✗
<i>Infomap</i>	✗	✗	✗	✗	✓	✓	✓	✓
<i>Label propagation</i>	✗	✗	✗	✗	✗	✗	✗	✗
<i>Eigenvector modularity</i>	✗	✗	✗	✗	✗	✗	✗	✗

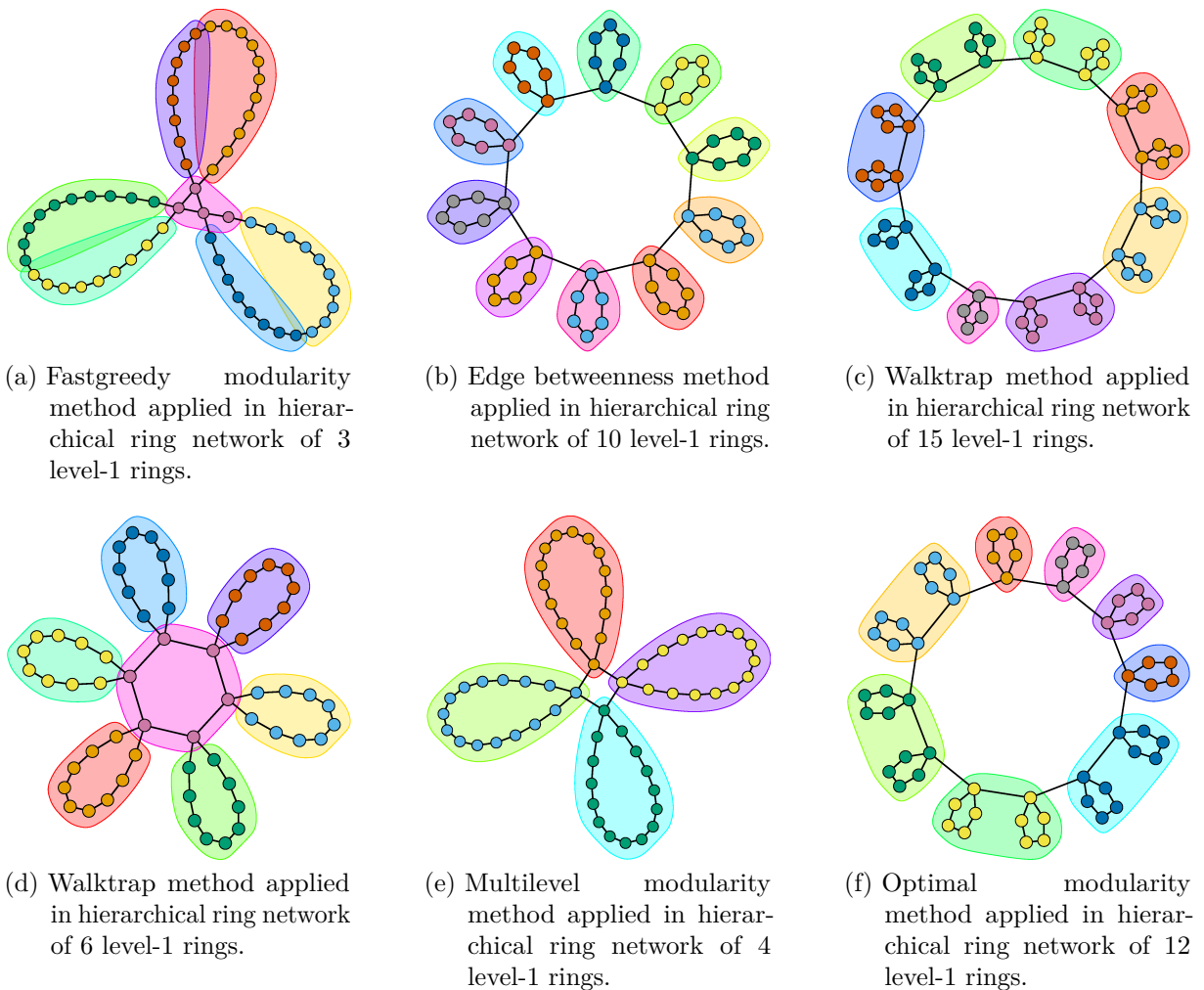


Figura 59 – Some results of community detection methods listed in Table 13.

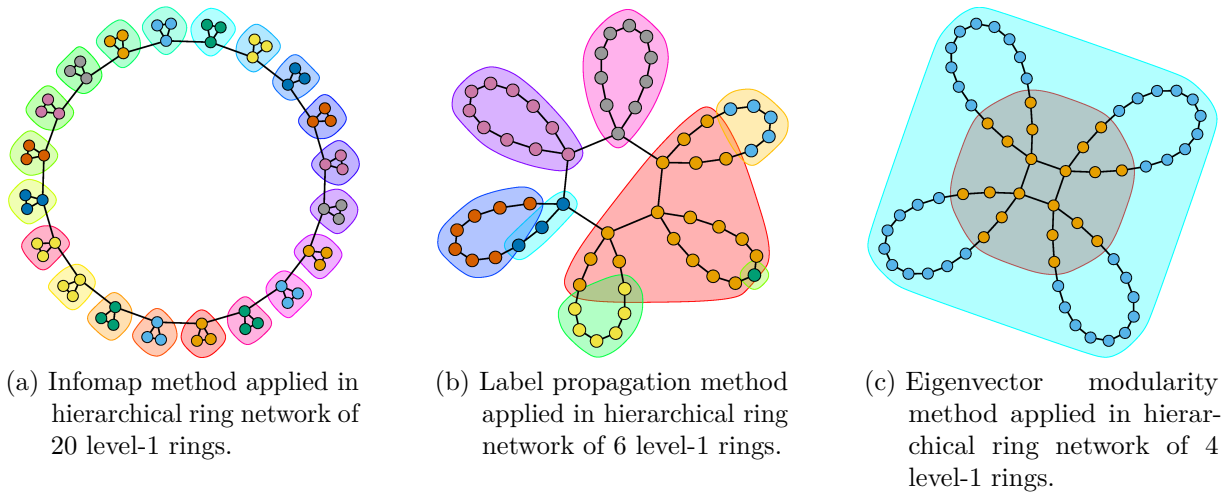


Figura 60 – Some results of community detection methods listed in Table 13.

The hierarchical ring network with 15 level-1 rings is the one that has most methods that respond in an approximate way, as illustrated for the method Walktrap in Figure 59c, where it can be observed that the communities are not identified in isolation, but rather in pairs. Since the number of rings is odd, then a ring is isolated.

In all networks, all detections that occurred as expected, occurred in the form shown in Figure 59b, that is, by identifying all level-1 rings as communities, without identifying the center level-0 ring, with only one exception. The exception occurred for the Walktrap method, and only in the network of 6 level-1 rings, as shown in Figure 59d.

Two methods are equally highlighted in the detection of the hierarchical rings: the

Tabela 14 – The values of the modularity index obtained for each hierarchical ring network analyzed, from the division of each network by each community detection method.

Community detection methods	Number of level-1 rings per graph							
	3	4	5	6	10	12	15	20
<i>Edge betweenness</i>	0,69	0,69	0,72	0,74	0,76	0,75	0,76	0,77
<i>Multilevel modularity</i>	0,69	0,69	0,72	0,74	0,76	0,75	0,76	0,78
<i>Optimal modularity</i>	0,71	0,70	0,72	0,74	0,76	0,75	0,76	0,78
<i>Walktrap</i>	0,71	0,70	0,67	0,67	0,76	0,61	0,76	0,77
<i>Spin-glass</i>	0,71	0,70	0,70	0,74	0,76	0,75	0,72	0,72
<i>Fastgreedy modularity</i>	0,71	0,68	0,69	0,72	0,76	0,75	0,76	0,77
<i>Infomap</i>	0,68	0,67	0,67	0,70	0,76	0,75	0,73	0,70
<i>Label propagation</i>	0,66	0,61	0,68	0,61	0,69	0,74	0,75	0,73
<i>Eigenvector modularity</i>	0,31	0,34	0,34	0,28	0,00	0,00	0,00	0,00
Minimum	0,31	0,34	0,34	0,28	0,00	0,00	0,00	0,00
Mean	0,65	0,64	0,66	0,66	0,67	0,65	0,67	0,67
Maximum	0,71	0,70	0,72	0,74	0,76	0,75	0,76	0,78

Edge betweenness and the Multilevel modularity, both identifying 5 networks as expected, and two in an approximate way, as can be seen in the first two lines of Table 13. In Figure 59b, we have already shown a successful case of the Edge betweenness method, and in Figure 59e we can see a successful case for the Multilevel modularity method, in this case in the network of 4 level-1 rings.

The Optimal modularity method performed a little worse, failing in two cases more than the two methods mentioned above. It is noted that both the Optimal modularity and Multilevel modularity methods are based on the same metric – the modularity index, and that they have obtained different performances, and this difference is explained by the way that each method applies the modularity index. In Figure 59f, we can see a case of failure for the Optimal modularity method.

The Walktrap method has fewer hits than Optimal modularity, finding the expected result only for networks with 6 and 10 level-1 rings. But, as it has already been said, the Walktrap method is the only one that considers the central ring (level-0 ring) as a community (Figure 59d). The Spin-glass method comes next, with a performance very similar to the Walktrap method.

The Fastgreedy modularity method performed very poorly, with only one expected result and another approximate result. It is worth noting that this method is based on the modularity index, in the same way as the Multilevel modularity method, which obtained good rate hits. Again, it is not enough to choose well the metric to be used in the community detection, once the way that the metric is applied can differ between being among the best or worst methods.

The Infomap method draws attention because it performs well as the number of rings increases, and consequently their sizes decrease. Therefore, it is inferred that this method can be indicated for the detection of smaller communities. The reason for this may be that the Infomap method uses minimization of the quantity of information needed to represent random walks, which we can suppose that it should be facilitated in graphs with smaller communities. An example of its application can be seen in Figure 60a.

Finally, the methods Label propagation and Eigenvector modularity presents the worst results, not identifying the rings in an expected way in any of the hierarchical ring networks analyzed, as it can be seen in Table 13. Figure 60b illustrates, for the Label propagation method, how much its results may not go well. The same can be seen in Figure 60c for the Eigenvector modularity method. It is interesting to note that both the best and the worst results are obtained by methods based on the modularity index.

Observing the unexpected results obtained, such as the one given by the Label propagation method (Figure 60b) – where, for example, an isolated node is considered a community within a ring – and still knowing that rings do not form very dense communities,

we raise the question of how the performance of this method could be with less trivial communities in more sophisticated networks.

A.3.2 Analysis of the modularity index against the detected communities

For each community detection method tested, and for each hierarchical ring network analyzed, the modularity index is calculated according to Equation A.1. The results of these calculations are displayed in Table 14. We verified that to present these data in graphical format does not seem advantageous, due to scale issues.

Let us first note that the values of modularity generated by the division of communities of the Eigenvector modularity method assume the smallest values, and also the most distant ones of all the other methods of Table 14, for all analyzed networks. It has already been observed in Table 13 that this method does not detect any community as expected, and therefore its modularity index values actually reflect this bad division. However, we also see in Table 13 that the Label propagation method got a result as poor as the Eigenvector modularity method's, but with modularity index values in Table 14 very close to the best results (such as Edge betweenness and Multilevel modularity). This shows that the modularity index does not reflect well the fact that community detection occurs or not as expected in the hierarchical networks considered here.

In Table 14, when comparing the results for the network with 3 level-1 rings (for which no method is successful in detecting the rings as communities) with the results for the case of the network with 10 level-1 rings (which has the largest number of methods identifying the rings as communities), and disregarding the Eigenvector modularity method, a very small difference in modularity indices is observed. A larger difference between the two groups did not occur, since the communities detected in the network of 3 level-1 rings actually have proportionally denser connections inside than outside them (for example, Figure 59a). Here again, it is observed that the modularity index does not reflect in a very sensitive way the fact that the method has or has not identified the rings in the expected way in the hierarchical ring networks.

A.4 Conclusion

In the present study, 9 community detection methods are applied in 8 hierarchical ring networks of 60 nodes each, in a controlled experiment.

The methods that best solved this issue, identifying such rings as network community structures, in most networks, are Edge betweenness and Multilevel modularity. On the other hand, the methods Label propagation and Eigenvector modularity did not reach such identification in any of the analyzed networks. It is remarkable that, based only on the method used, the modularity index is able to be part of the best and worst method

for the identification of rings in hierarchical ring networks, where what is varying between the two methods is only how the index is optimized.

Analyzing the computations of the modularity index for all networks, based on the division into communities performed by each method, we identified that this measure has a very weak sensitivity to inform whether or not the division in communities, generated by the method, has identified the rings in hierarchical ring networks.

The obtained results indicate the need of questioning what they actually detect and how to evaluate their results when applying algorithms to detect communities in trivial or complex networks. This questioning is essential for the correct interpretation of the results provided by the algorithms.

We ran preliminary tests, using full subgraphs instead of rings, and the results are also unsatisfactory. This research will be further explored in future studies. Hierarchical ring networks with rings at levels greater than one can also be tested. But, since our goal here is not to detect rings in hierarchical networks, but rather to question how well the methods of detecting communities work in the role they play, we also have as future work an analysis, from the point of view of algorithms, to identify strengths which can lead to an improvement in the quality of detection of communities in general.

Lista de Abreviaturas e Siglas

NSFNET	<i>National Science Foundation Network</i>
OTN	<i>Optical Transport Networks</i>
WDM	<i>Wavelength Division Multiplexing</i>
WRON	<i>Wavelength Routed Optical Networks</i>
RWA	<i>Routing and Wavelength Assignment</i>
ILP	<i>Integer Linear Programming</i>
EON	<i>Elastic Optical Networks</i>
OOFDM	<i>Optical Orthogonal Frequency Division Multiplexing</i>
RSA	<i>Routing and Spectrum Assignment</i>
CAPEX	<i>Capital Expenditure</i>

Lista de Símbolos

λ	Número de Comprimentos de Onda
G	Grafo
V	Conjunto de vértices
E	Conjunto de arestas
$G(V,E)$	Grafo formado pelos vértices de V e pelas arestas de E
n	Número de vértices do grafo, ou ainda a ordem da rede ($ V = n$)
m	Número de arestas do grafo, ou ainda tamanho da rede ($ E = m$)
α	Densidade de aresta
C	Cardinalidade
H	Entropia quadrática
Δ	Vizinhança de coincidência
C_R	Cardinalidade relativa
$I(X,Y)$	Informação mútua entre X e Y

Lista de Tabelas

1	Alguns parâmetros básicos das 15 redes do mundo real analisadas. . . .	36
2	Os 30 invariantes que reduzem a pelo menos a metade a incerteza do número de comprimentos de onda, para as $2, 2 \times 10^6 - 52$ redes aleatórias analisadas. Esses invariantes são ordenados em função de $I(\lambda; k)$ médio, e a cada um é atribuído um identificador “k” dado em função dessa ordenação. O Apêndice B exibe uma tabela com todos os 315 invariantes em estudo, ordenados por $I(\lambda; k)$ médio.	42
3	Os 30 invariantes mais importantes de acordo com os valores de $I(\lambda; k)$, para as $2, 2 \times 10^6 - 52$ redes aleatórias analisadas, separadas pela ordem das redes. Células em branco indicam invariantes presentes em todas as colunas. Por outro lado, células coloridas indicam invariantes que não estão presentes em pelo menos uma coluna. O Apêndice C apresenta esta tabela de forma mais extensa, para o total dos 315 invariantes estudados.	47
4	Matriz de correlação linear dos 14 invariantes mais representativos, i.e., aqueles que aparecem na Tabela 2 e em todas as colunas da Tabela 3. .	48
5	Invariantes mais relevantes para explicar o número de comprimentos de onda (k é o identificador do invariante). O apêndice A exibe uma breve descrição e o modo de cálculo desses invariantes.	58
6	Resultados das diferenças entre cada λ e seus valores estimados.	68
7	Comparação entre λ , $E[\lambda]$ e $E_{min}[\lambda]$	77
8	Valores de λ das redes reais de entrada versus os valores de λ das redes de saída (G_S) das heurísticas, para cada vizinhança possível. As células preenchidas de rosa são aquelas que guardam o valor mínimo encontrado por rede dentre todas heurísticas.	79
9	Detalhamento dos resultados das redes reais para a heurística VNS-KDS.	82
10	Valores de S_2 , S_3 e S_4 para todas as redes reais em análise.	85
11	Valores de $R_G(\rho)$ para todas as redes reais em análise e suas respectivas redes de saída.	86
13	Results of the application of the community detection methods, for each hierarchical ring network analyzed. The symbols ✓ and ✗ indicate, respectively, success and failure in rings detection. An +/- indicates that the detection of the rings is not as expected, but that the result has some coherence.	165

14	The values of the modularity index obtained for each hierarchical ring network analyzed, from the division of each network by each community detection method.	166
----	--	-----

Lista de Figuras

1	Ilustração representativa da extensa rede de comunicação digital global. Fonte: https://pngtree.com (2018).	12
2	Em (a), tem-se uma ilustração da interligação da rede óptica <i>backbone</i> NSFNET com pontos de redes regionais menores, e em (b) apenas a topologia física da referida rede <i>backbone</i>	13
3	Ilustração da fragmentação espacial e da fragmentação espectral. A barra da esquerda representa o espectro do enlace que liga A a B, e a barra da direita representa o espectro do enlace que liga B a C. Ao buscar atender uma demanda de tráfego de 4 <i>slots</i> de A até C, ocorre a fragmentação espacial, pois não há disponibilidade contínua de 4 <i>slots</i> em cada fibra, e ocorre ainda a fragmentação espectral, pois de A a B não há 4 <i>slots</i> contíguos (Figura adaptada de Izquierdo-Zaragoza, Pavon-Marino e Bueno-Delgado (2014) <i>apud</i> Tessinari (2016)).	14
4	Em (a) e em (b), três canais são transmitidos, onde em (a) essa alocação utiliza três comprimentos de onda, e já em (b) utiliza dois.	15
5	Duas topologias de rede com o mesmo número de nós ($n = 10$) e enlaces ($m = 15$). A topologia (a) requer 8 comprimentos de onda, enquanto a topologia (b) requer 12 comprimentos de onda, i.e., 50% mais comprimentos de onda que a topologia (a).	24
6	Diagrama que resume delimitação do tema da presente tese.	26
7	Visualização do perfil gerado pela ordenação de todos os valores de informação mútua dos 315 invariantes analisados.	41
8	Visualização do perfil gerado pela ordenação de todos os valores de informação mútua dos 30 invariantes que atenderam o critério de $I(\lambda;k) \geq 2$, ou seja, de reduzir a incerteza com relação à λ a, no mínimo, a metade.	42
9	<i>Edge betweenness</i> máximo ($k = 1$) versus λ para as redes reais (representadas por losangos pretos) e redes aleatórias em estudo: para todas as ordens de rede (a), e separadas por n (b). Em (c), informação mútua do <i>Edge betweenness</i> máximo e λ , em função da ordem da rede.	44
10	O número de vértices ($k = 28$) versus λ , para todas as redes reais e aleatórias em estudo. As cores representam o grau médio, e o número em cada losango é o grau médio das redes do mundo real estudadas.	45
11	Comparação dos invariantes que foram mais importantes na análise com todas as redes com ordens n juntas e separadas.	50

12	Coeficiente de variação da distância de comunicabilidade ajustada ($k = 12$) versus λ para todas as redes do mundo real (representadas por losangos) e redes aleatórias em estudo: redes com todas as ordens n juntas em (a), e separadas pela ordem n em (b). Em (c), é dada a informação mútua do coeficiente de variação da distância de comunicabilidade ajustada e λ , em função da ordem da rede.	52
13	Índice de Harary ($k = 14$) versus λ para todas as redes do mundo real (representadas por losangos) e redes aleatórias em estudo: redes com todas as ordens n juntas em (a), e separadas pela ordem n em (b). Em (c), é dada a informação mútua do Índice de Harary e λ , em função da ordem da rede.	52
14	Informação Mútua da (a) distância média ($k = 7$) e da (b) transmissão mediana ($k = 11$), ambos com λ e em função da ordem da rede (n). . .	53
15	Densidade de arestas ($k = 65$) e desvio padrão do grau ($k = 211$) ambos versus λ , para as redes do mundo real (representadas por losangos) e redes aleatórias em estudo: para todas as redes de todas as ordens n juntas em (a) e (c), e separadas pela ordem n em (b) e (d).	54
16	Densidade de arestas ($k = 65$) versus λ para redes aleatórias em estudo de todas as ordens n juntas.	55
17	Grau médio ($k = 167$), conectividade algébrica ($k = 40$) e distância média ($k = 7$), todas versus λ para todas as redes do mundo real (representadas por losangos pretos) e redes aleatórias em estudo: para redes de todas as ordens n em (a), (c), e (e), e para redes separadas por n em (b), (d), e (f).	57
18	Etapas do processo para obtenção de topologias com baixo requerimento de comprimentos de onda (λ), baseado numa otimização que usa funções substitutas.	62
19	Histograma de λ das redes de S.	63
20	Tempo de CPU versus o número de nós n para o cenário de teste, o qual compara o método de estimação de λ via as Equações 3.3 e 3.4 (em verde) e a solução exata (em vermelho). Dados de tempo computacional gasto na solução exata obtidos em Cousineau <i>et al.</i> (2015).	66
21	Visualização da relação linear (em vermelho) entre $E[\lambda]$ e λ em (a), e entre $E_{min}[\lambda]$ e λ em (b). Dados da segunda amostra de 3300 topologias (em azul) em ambos (a) e (b).	67
22	Heurística baseada na busca aleatória, ou seja, na escolha randômica das topologias a serem testadas.	69
23	Heurística de busca local do tipo <i>Downhill</i> (algoritmo guloso).	71
24	Heurística da busca local com inspeção completa das vizinhanças. . . .	73

25	Heurística baseada na busca em vizinhança variável.	75
26	Cenário 1: Dois casos da NSFNET com um enlace adicionado em vermelho tracejado. Em (a), $\lambda = 11$ (melhor caso) alcançado em 18 das 70 possibilidades, e em (b), $\lambda = 14$ (pior caso) alcançado em 6 das 70 possibilidades.	87
27	Em (a), topologia real da NSFNET, com $n = 14$, e $m = 21$, a qual tem $\lambda = 13$. Em (b), Cenário 2: Topologia gerada com $n = 14$, e $m = 22$, a qual possui $\lambda = 9$	88
28	Maximum of edge betweenness ($k = 1$) and Standard deviation of edge betweenness ($k = 2$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum of edge betweenness and SD of edge betweenness both with λ , as a function of network order.	123
29	Amplitude of edge betweenness ($k = 3$) and Kirchhoff index ($k = 4$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Amplitude of edge betweenness and Kirchhoff index both with λ , as a function of network order.	124
30	Mean of edge betweenness ($k = 5$) and Mean of vertex betweenness ($k = 6$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean of edge betweenness and Mean of vertex betweenness both with λ , as a function of network order.	125
31	Mean distance ($k = 7$) and Mean transmission ($k = 8$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean distance and Mean transmission both with λ , as a function of network order.	126
32	Maximum transmission ($k = 9$) and Wiener index ($k = 10$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum transmission and Wiener index both with λ , as a function of network order.	127

33	Median transmission ($k = 11$) and Coefficient of variation of adjusted communicability distance ($k = 12$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Median transmission and Coefficient of variation of adjusted communicability distance both with λ , as a function of network order.	128
34	Median of edge betweenness ($k = 13$) and Harary index ($k = 14$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Median of edge betweenness and Harary index both with λ , as a function of network order.	129
35	Standard deviation of transmission ($k = 15$) and Mean eccentricity ($k = 16$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Standard deviation of transmission and Mean eccentricity both with λ , as a function of network order.	130
36	Amplitude of transmission ($k = 17$) and Coefficient of variation of modularity ($k = 18$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Amplitude of transmission and Coefficient of variation of modularity both with λ , as a function of network order.	131
37	Maximum of adjusted betweenness centrality ($k = 19$) and Minimum transmission ($k = 20$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum of adjusted betweenness centrality and Minimum transmission both with λ , as a function of network order.	132
38	Standard deviation of vertex betweenness ($k = 21$) and Standard deviation of adjusted betweenness centrality ($k = 22$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Standard deviation of vertex betweenness and Standard deviation of adjusted betweenness centrality both with λ , as a function of network order.	133

39	Maximum of vertex betweenness and Vertex betweenness centrality both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Maximum of vertex betweenness and Vertex betweenness centrality both with λ , as a function of network order.	134
40	Mean of adjusted communicability distance ($k = 25$) and Amplitude of vertex betweenness ($k = 26$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean of adjusted communicability distance and Amplitude of vertex betweenness both with λ , as a function of network order.	135
41	Amplitude of adjusted betweenness centrality ($k = 27$) and Vertices number ($k = 28$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Amplitude of adjusted betweenness centrality and Vertices number both with λ , as a function of network order.	136
42	Interquartile distance of edge betweenness ($k = 29$) and Standard deviation by median of edge betweenness ($k = 30$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Interquartile distance of edge betweenness and Standard deviation by median of edge betweenness both with λ , as a function of network order.	137
43	Coefficient of variation of cocitation coupling ($k = 33$) and Coefficient of variation of dice similarity ($k = 34$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of cocitation coupling and Coefficient of variation of dice similarity both with λ , as a function of network order.	138

44	Coefficient of variation of inverse log-weighted similarity ($k = 35$) and Mean of inverse log-weighted similarity ($k = 36$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of inverse log-weighted similarity and Mean of inverse log-weighted similarity both with λ , as a function of network order.	139
45	Algebraic connectivity ($k = 40$) and Standard deviation of distance ($k = 45$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Algebraic connectivity and Standard deviation of distance both with λ , as a function of network order.	140
46	Mean of cocitation coupling ($k = 46$) and Edges number ($k = 52$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Mean of cocitation coupling and Edges number both with λ , as a function of network order.	141
47	Degree centralization ($k = 53$) and Median eccentricity ($k = 55$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Degree centralization and Median eccentricity both with λ , as a function of network order.	142
48	Coefficient of variation by median of dice similarity ($k = 57$) and Number of minimum sets that disconnects the graph ($k = 59$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation by median of dice similarity and Number of minimum sets that disconnects the graph both with λ , as a function of network order.	143
49	Coefficient of variation of eccentricity ($k = 60$) and Coefficient of variation by median of Jaccard similarity ($k = 61$) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of eccentricity and Coefficient of variation by median of Jaccard similarity both with λ , as a function of network order.	144

50	Coefficient of variation by median of inverse log-weighted similarity (k = 63) and Coefficient of variation by median of cocitation coupling (k = 68) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation by median of inverse log-weighted similarity and Coefficient of variation by median of cocitation coupling both with λ , as a function of network order.	145
51	Coefficient of variation of edge betweenness (k = 93) and Minimum of subgraph centrality (k = 96) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of edge betweenness and Minimum of subgraph centrality both with λ , as a function of network order.	146
52	Minimum of average nearest neighbor degree (k = 101) and Number of spanning trees (k = 109) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Minimum of average nearest neighbor degree and Number of spanning trees both with λ , as a function of network order.	147
53	Coefficient of variation of distance (k = 115) and Coefficient of variation by median of transmission (k = 126) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation of distance and Coefficient of variation by median of transmission both with λ , as a function of network order.	148
54	Coefficient of variation by median of closeness centrality (k = 138) and Coefficient of variation by median of closeness centrality normalized (k = 139) both versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a) and (b), and separated by n in (c) and (d). In (e) and (f), mutual information of Coefficient of variation by median of closeness centrality and Coefficient of variation by median of closeness centrality normalized both with λ , as a function of network order.	149
55	Mean of vertex degree (k = 167) versus λ for all real-world (represented by lozenges) and random networks under study: for all network orders in (a), and separated by n in (b). In (c), mutual information of Mean of vertex degree and λ , as a function of network order.	150

56	Network (a) is formed by joining 3 complete graphs of 20 nodes each, and network (b) is formed by joining 3 cycles of 20 nodes each. In both networks the Edge betweenness method is used for the division into communities. In (a), the modularity value is 0.66, and in (b) is 0.69, even (a) presenting greater visual structuring of communities than (b).	161
57	Example of a hierarchical ring network, with the center ring of level 0 (of 6 nodes) binded to others 3 rings of level 1.	161
58	Hierarchical ring networks of 60 nodes divided homogeneously.	162
59	Some results of community detection methods listed in Table 13.	165
60	Some results of community detection methods listed in Table 13.	166